

Policing ourselves

Biologists should push forward with an effort that began in California last weekend to wrestle with the implications of synthetic biology.

There are many ways of measuring the importance of a scientific discipline, and most would lead you to dismiss synthetic biology as a pretty marginal affair. It has yet to produce a profusion of great papers, to underpin vibrant new businesses, to generate stereotypes in the public imagination, to benefit from vast flows of grant money, or to boast a substantial body of practitioners united in their vision for their field. And a large element of the community can fit into a small lecture theatre, as it did last week at the Synthetic Biology 2.0 conference at the University of California, Berkeley (see page 388).

But some indicators speak differently. To participate in the meeting was to witness an inspiring interdisciplinary festival of ideas — and the central one of supplementing biology's attempts to understand 'life as it is' with systematic explorations of 'life as it could be' has undeniable élan. The field has attracted venture capitalists, Nobel laureates, acres of newsprint and Craig Venter (although there are some who would dispute whether any or all of these correlate with true worth).

Synthetic biology is also important enough to have attracted enmity. The Berkeley meeting was greeted by an open letter from 35 groups, claiming standing in 60 countries, that denounced the ambitions of synthetic biology and objected to the idea that its practitioners might institute structures of self-governance to mitigate some of its inherent risks — an idea mooted before the meeting, but one that remains an aspiration rather than a fixed intention.

The most powerful of the field's intellectual attractions is its bottom-up nature. Previous attempts to alter or improve on nature have been top-down, starting off with a whole, functioning organism and tweaking it, whether through selective breeding, by the addition of transgenes, or with a vaccine to focus the immune system. But in the light of the emerging technological possibilities of gene synthesis, synthetic biology aspires to work from the bottom up, building things from scratch, to remake, rather than reshape, the world, starting with a blank sheet of paper.

Blank sheets of paper are, for most of us, rather scary. The currently ill-defined and potentially immense capacities of synthetic

biology bring worries aplenty, from unintended consequences to deliberate malfeasance. One does not have to agree with the letter's authors to think that, despite offering the possibilities of great benefit, synthetic biology also raises some very significant concerns.

So although many of the specific issues raised in the open letter fail to compel agreement, the overall message that there are real concerns is a valid one. What seems less defensible is the letter's hostility to self-governance; its authors are, after all, keen that the issues be discussed, and implicitly that governance be exercised.

Self-governance need not and should not be exclusive — it does not preclude other forms of governance, any more than the possession of conscience makes redundant the strictures of law. It is hard not to suspect that the problem with self-governance from the point of view of the letter-writers is that it could go some way to addressing potential problems that would make good campaigning issues.

The ability of human societies to modify and transform biological systems will increase more in this century than it has in the hundred centuries since the dawn of agriculture, regardless of whether the transformation unfolds under the rubric of 'synthetic biology'. Or, at least, we must hope that it will — as the only credible alternative is a future in which massive social upheaval, armed conflict or natural disaster halts the progress of scientific knowledge. The challenge is to foster a matching, or at least sufficient, increase in the wisdom and accountability with which these abilities are used.

That challenge will require changes in the law and customs, in ideology and theology, and in education and economics. No scientific community can be expected to shoulder all that on its own, and nor should it. Scientists who are alive to the possibilities of change, anxious to keep their house in order and be seen to be doing so, and keen to discuss the issues with the world, are part of the solution, not part of the problem. ■

"Scientists who are keen to discuss the issues with the world are part of the solution, not part of the problem."

Coping with complexity

A more detailed understanding of scientific concepts does not lead to simplicity.

Two philosophers of science recently surveyed 500 geneticists to ask their opinion on whether 14 different sets of genetic information constituted a gene, or more than one gene. Fortunately, the bulk of the respondents felt able to answer the questions

definitively. Less fortunately, their answers were inconsistent, with the sample often quite evenly split on the question of how many genes were actually present.

Sceptics might note a degree of unravelling here. Decades of discussion have left a rather widespread perception, embraced by the general public and the media, of the gene as a tightly defined entity that spells out an inescapable destiny filled with beauty and health or, more often, blemishes and disease.

Even among the medical profession and some scientists, a gene is a trusty and well-defined concept — a specific sequence of genetic



information that, when converted into messenger RNA, encodes a protein. Following this line of thinking, all that medical researchers really need to do is link up those darned diseases with their underlying genes, and human biology would fall into place.

Among geneticists themselves, this notion has long been eclipsed. Where once scientists saw placid, lonely genes that mass-produce RNA transcripts, now they find a chaotic jumble of RNA generated from all over the genome and from outside conventional genes. They have little clue what this RNA is doing, and don't always know where one gene ends and the next begins (see page 399).

And anyway, forget about DNA, says a paper on page 469 of this issue; some RNA might also be ferrying information from one generation to the next.

For most geneticists, this complexity is a source of marvel and fascination — and employment. How dull their lives would be if, once the human genome had been sequenced, there were just genes and diseases to be linked, like one of those join-the-dots puzzles. The genetic code holds new allure — its four-letter sequence may have been documented but it contains deeper hidden ciphers, and geneticists relish the task of breaking them.

There remains a nagging concern, however, that some of these challenges will frustrate the hopes of earlier generations that the

study of biology could reduce complex problems to a mechanistic understanding of the relationship between DNA and living things.

Indeed, the complexity of this relationship has got us to the point where geneticists find it hard to agree on an appropriate definition of a gene. They are also unsure whether genes themselves are worthy of the most attention, compared with other parts of the genome, or RNA or proteins, or the way they all interact together in different tissues. At the very least, a serious disconnect seems to have arisen between the real problems that geneticists are wrestling with, and the public understanding of what they do.

It falls on researchers to make sure that the gap doesn't grow too wide. That means conveying the complexity of the task more clearly and fighting the media's tendency to boil down complex investigation to the 'discovery of the gene for something'. Geneticists should not be afraid of using new words, such as transcripts or loci, if these serve them better and more accurately. The public can cope with such distinctions.

After all, 'gene' is just a word, and dictionaries can be revised. Bring on the complexity — biology would be boring without it. ■

"The genetic code holds new allure — its four-letter sequence may have been documented but it contains deeper hidden ciphers."

Carbon omissions

The European Union's greenhouse-gas trading system needs reinforcement.

When the European Union (EU) started the world's first mandatory greenhouse-gas emissions trading system in January 2005, there was widespread scepticism about its prospects. Thousands of companies were allocated 'allowances' for emitting a given amount of carbon dioxide; if they produce more, they have to buy extra credits. The sceptics questioned whether prices would climb high enough to provide a real incentive for industry to invest in clean technologies, or for power producers to switch to less-carbon-rich energy sources.

Some of this criticism subsided when the price, as well as the volume of trading on the new market, began to increase. By the middle of last month, the price of credits to release a tonne of carbon dioxide had reached more than €30 (US\$39). Then the price fell sharply, on rumours of unexpectedly low industrial emissions in a number of EU countries. But although the European Commission officially confirmed on 15 May that EU industries emitted 44 million tonnes less carbon dioxide in 2005 than they had been allocated, the predicted market meltdown failed to materialize (see page 405).

There are lessons to be learned from this month's market turmoil. Clearly, the European Commission, which is in charge of approving each member state's emissions allocation, over-allocated such rights for the initial trading period, from 2004 to 2007. But that is no reason to fear for the market's survival — it is normal, in the history of using such markets to control pollutants, for regulators to aim low and then turn up the heat as more experience is gained.

It is equally clear that administrative questions regarding the

estimation, allocation and verification of emissions have yet to be fully resolved. Reporting and verification methods vary from country to country, with some, including Poland and Italy, admitting that their national registries aren't yet functioning.

Additionally, the basis of allocating emissions rights to different sections of industry isn't sufficiently clear or logical, and can actually hurt growth sectors that have actively sought to cut emissions but whose efforts are not rewarded by the system. A British-led initiative to update the allocation system has unfortunately come to nothing.

These issues need to be addressed if emissions trading is to remain viable and reach beyond the EU. A review of technological improvements in pollution control made since 1990 — the baseline year for the allocation assessment — is needed, along with greater harmonization of the allocation and verification processes. Independent inspectors accredited by the European Commission must be granted full access to facilities across the EU.

Ultimately, the market will have to be transparent, and the price substantial and reasonably stable, if industry is to take it into consideration when making investment decisions. A power company will not spend hundreds of millions of euros on carbon-sequestration facilities unless prices on the market signal clearly that it is economically sensible to do so. As such investments can take 20 years to bear fruit, industry also needs to know the system will still be around for the long haul.

The EU will in due course provide these assurances. The European Commission, in the meantime, must press on with the task of creating a firm and fair regulatory framework within which the market can thrive. ■

"The market will have to be transparent, and the price substantial and reasonably stable, if industry is to take it into consideration when making investment decisions."

RESEARCH HIGHLIGHTS

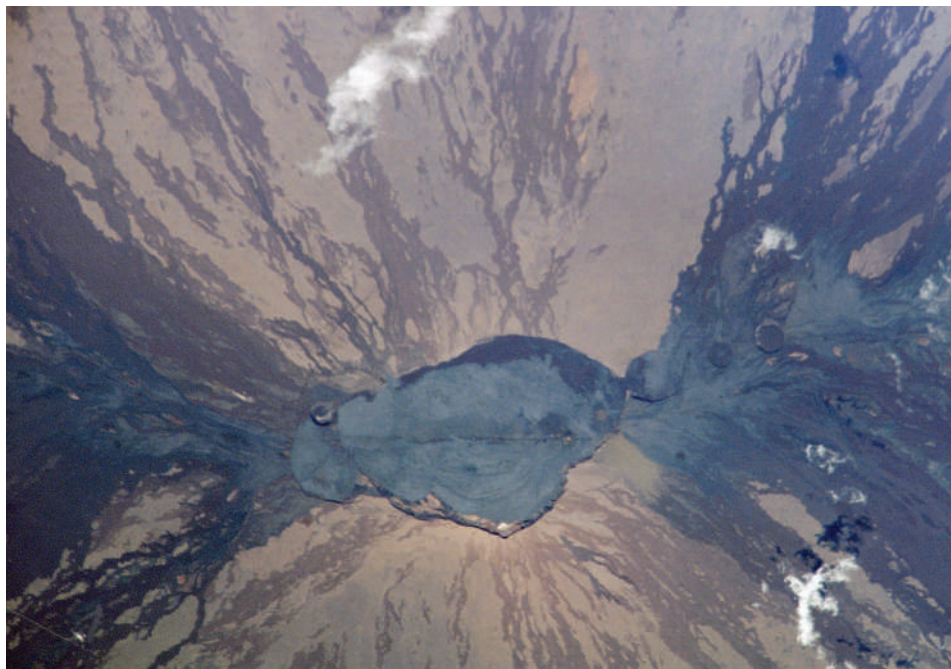
A bigger bang

Geochim. Geophys. Geosyst. doi:10.1029/2005GC001086 (2006)

Mauna Loa (pictured), the world's largest volcano, is even more enormous than thought, a recent study suggests.

Submersible dives and seafloor mapping studies yielded a new underwater map of Mauna Loa's western flank, including ten submarine vents where only one was known before. The data suggest that its 1877 underwater eruption, which made the surface of a nearby bay boil, was much larger than thought.

If so, the volume of Mauna Loa's historical eruptions is 10% greater than previously believed, contradicting the theory that this Hawaiian volcano is dying, say Dorsey Wanless of the University of Hawaii in Honolulu and colleagues. The volcano last erupted in 1984.



NASA/SPL

CIRCADIAN BIOLOGY**Periodic regulation**

Mol. Cell 22, 375–382 (2006)

The circadian feedback loop known as the biological clock has recently been indirectly implicated in cancer development.

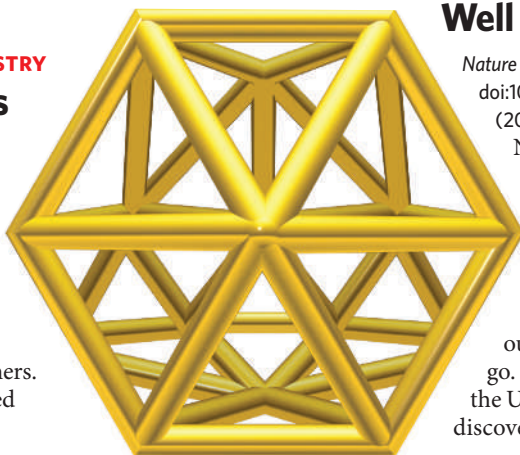
Researchers have now provided more direct evidence, by showing that a gene called *Per1*, a key component of the circadian clock, regulates growth and DNA damage in human cancer cells.

A team led by Sigal Gery at the University of California, Los Angeles, reports that overexpression of *Per1* interferes with the cell's complex molecular defences against the cancer-inducing effects of ionizing radiation. What's more, the group found reduced levels of *Per1* in human cancer patient samples.

Theory predicted that a few dozen gold atoms could naturally order themselves into a structure similar to carbon-60 'buckyballs', but early experiments showed that groupings of 32 or more atoms resembled little more than a tangled gold nugget.

Now Lai-Sheng Wang of Washington State University, Richland, Xiao Cheng Zeng of the University of Nebraska, Lincoln, and their team have examined clusters of 20 gold atoms or less and found a wide range of naturally occurring arrangements — from pyramids to gold cages (pictured below).

The authors believe the cages could be used to hold single atoms of various types, potentially useful in nanotechnology applications.

**CELL BIOLOGY****Well sorted**

Nature Struct. Mol. Biol.
doi:10.1038/nsmb1098
(2006)

Newly made proteins need to find their way to the right places in the cell.

To do this, they carry a molecular address label spelling out where they should go. Now, a team in the United States has discovered a molecular

courier that reads labels and helps to deliver proteins to an exclusive location in the cell's nucleus.

Sharon Braunagel of Texas A&M University and her colleagues studied proteins destined for the inner of the two membranes bounding the nucleus. They discovered another protein, called importin- α -16, that identifies these molecules and delivers them, possibly by hitching a lift on one of the cell's motor proteins.

PLANETARY SCIENCE**Magnetic sunscreen**

Astron. Astrophys. 451, L43–L46 (2006)

An asteroid's colour may reveal whether it has a magnetic field to shield it from the ageing effects of the solar wind.

Space rocks are darkened and reddened by the stream of ions from the Sun. But the surface of Vesta, the second-largest asteroid known in our Solar System, is surprisingly pristine.

Simulating this process in the lab, Pierre Vernazza of the Paris Observatory, France, and colleagues exposed a meteorite, thought to originate from Vesta, to ions in the lab to show that its parent should indeed be substantially more weathered than it appears. They suggest that the asteroid must have a magnetic field of at least 0.2 microtesla at its surface, a few hundred times smaller than Earth's own field, which diverts the damaging ions.

PNL

PHYSICAL CHEMISTRY**Goldenballs**

Proc. Natl Acad. Sci.
doi:10.1073/pnas.
0600637103 (2006)

Forming football-shaped molecules is no longer the sole preserve of carbon atoms, say researchers. Now gold has joined the game.

ANIMAL BIOLOGY

Good shot

Curr. Biol. **16**, R316–R318 (2006)

Jellyfish stings can puncture their targets with the power of rifle bullets, say researchers in Germany. The study reveals intriguing insight into how these tiny structures, known as cnidae, penetrate the tough skins of prey.

Thomas Holstein of the University of Heidelberg and his colleagues used an ultra-high-speed ballistics camera to watch stinging cells from the jellyfish-like animal *Hydra* (pictured right) discharge. They calculated that the sharply pointed stings were released at speeds of up to 40 metres per second (140 kph), with an acceleration of 5.4×10^6 g, resulting in a pressure on impact of 7 gigapascals.

This movement, one of nature's fastest, owes its force to powerful spring-like proteins that contract when the pressure in the cnidae drops as they discharge.



side of the neuronal synapse and helps to inhibit the release of neurotransmitters, whereas GABA_{B1b} localizes to the opposite side, where it helps to dampen neuronal firing of post-synaptic cells.

NEUROBIOLOGY

Same but different

Neuron **50**, 589–601; 603–616 (2006)

GABA (γ -aminobutyric acid), the main inhibitory chemical in the vertebrate nervous system, has many different modes of action. This is partly thanks to the precise location and activity of one of its receptor components, two new studies reveal.

The GABA_B receptor is made of two pieces, and one piece known as GABA_{B1} comes in two versions: GABA_{B1a} and GABA_{B1b}. Teams in Switzerland led by Bernhard Bettler of the University of Basel and Matthew Larkum of the University of Bern studied mice in which the different GABA_{B1} versions were inactivated one at a time.

They found that GABA_{B1a} localizes to one

MICROBIOLOGY

Stripped bare

PLoS Pathogens **2**, e35 (2006)

Researchers in the United States have uncovered a network of genes that shields a fungus' cell wall from immune-system attack.

In disease-causing fungi, an outer coat protects an inner layer of a sugary protein called β -glucan. Robert Wheeler and Gerald Fink of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, screened a library of 4,800 mutant yeast strains to find genes that, when disabled, expose the β -glucan. They showed that disabling nearly 50 of these genes makes fungi more vulnerable to attack by mouse immune cells.

Targeting the proteins made by the masking genes might yield much-needed medicines to fight insidious pathogenic fungi such as *Candida albicans*.

MATERIALS

Graphite unzipped

Phys. Rev. Lett. **96**, 176101 (2006)

Graphene — a single graphite-like carbon sheet — has unusual and potentially useful electronic behaviour. But it is hard to make. One promising approach involves oxidizing graphite to form graphite oxide, which separates layer by layer into thin flakes.

These flakes are generally much smaller than the original layers, and now Je-Luen Li and colleagues at Princeton University, New Jersey, think they know why.

The team identified bright lines in microscopic images of graphite oxide as fault lines where the layers have cracked. Their quantum-chemical calculations showed that these cracks are created by an 'unzipping' process in which the cooperative bonding of oxygen atoms next to one another on the surface creates a row of broken carbon-carbon bonds.

CHEMISTRY

Clean sweep

Science **312**, 1024–1026 (2006)

Hydrogen atoms can now be stripped from a silicon surface using laser light in a process that generates virtually no heat. Hydrogen desorption is an important part of computer-chip manufacture that normally requires temperatures of around 800 °C, which can introduce defects into the chips.

The technique, developed by Philip Cohen from the University of Minnesota, Minneapolis, and his colleagues, is highly selective, picking off hydrogen but leaving atoms of its heavier isotope deuterium behind. This shows that there is no transfer of heat across the silicon surface, as this would cause both hydrogen isotopes to desorb.

JOURNAL CLUB

Catherine L. Drennan
Massachusetts Institute of
Technology, USA

A biochemist considers cool inorganic chemistry that is relevant to our energy crisis.

There is a class of enzymes, found in certain microorganisms, that are the envy of synthetic chemists. They catalyse the reversible conversion of carbon dioxide (CO₂) to carbon monoxide (CO) — a

difficult feat in the lab because the carbon-oxygen double bond in CO₂ is very strong.

In striving to match nature's success, chemists seek twofold environmental benefits: making CO — a useful feedstock for synthesizing carbon compounds — without using precious fossil fuels while, at the same time, consuming the greenhouse gas CO₂.

After my group solved the crystal structure for two of these enzymes, known as carbon monoxide dehydrogenase/acetyl-CoA synthase, I sought out the

company of Joseph Sadighi, an assistant professor in my department, to discuss possible mechanisms involving the use of the enzymes' nickel-iron-sulphur clusters as catalysts. Sadighi was searching for other first-row transition metals that could catalyse the same reaction.

As my research took me in new directions, our long conversations about CO₂ came to an end. I heard occasional updates on his progress, but it wasn't until last October that I learnt the full story. I received a preprint of a paper

from Sadighi and his co-workers announcing their new catalyst for CO₂ reduction (D. S. Laitar, P. Müller & J. P. Sadighi *J. Am. Chem. Soc.* **127**, 17196–17197; 2005). Instead of nature's choice of iron and nickel, they had prepared copper(I) boryl complexes. These catalysts convert an impressive 100 molecules of CO₂ per hour at room temperature.

This is not only an exciting development in catalysis, it also shows the potential of inorganic chemistry in tackling our environmental problems.

NEWS

PHOTO TAKE



Peer review: biologists' growing ability to engineer life from its basic building blocks has raised concerns among biotech activists.

Synthetic biologists try to calm fears

BERKELEY, CALIFORNIA

Researchers in the field of synthetic biology are to issue a declaration of intent about professional behaviour and organization in order to ensure good practice and to address a range of concerns about their research. The scientists hope to ensure that controversy doesn't choke the field just as it begins to make progress. But critics are likely to be unimpressed — a coalition of organizations concerned about the technology released an open letter ahead of the meeting calling for the field to be externally regulated.

The scientists and engineers meeting this week in Berkeley for the Synthetic Biology 2.0 conference discussed recent advances towards their goal of being able to develop biological systems from scratch. But they also devoted a day of the meeting to issues of intellectual property, biosecurity, risks and standards. As *Nature* went to press, participants were drafting a set of statements to be posted online (<http://pbd.lbl.gov/sbconf>) for others to comment on. They intend to present

this as an outcome of the meeting.

One suggestion discussed at the conference was a possible commitment to buy synthesized DNA only from companies that screen orders for safety and security concerns, for instance by checking them against a list of dangerous pathogens. The community hopes this would force companies that do not already use such methods to adopt them, making it harder for those with malicious intentions to obtain deadly DNA.

"Researchers hope that by regulating themselves they can stave off attempts to set limits on the field."

The synthetic biologists hope that, by regulating themselves, they can stave off attempts to set controls or limits on the field. The science is moving rapidly: at the meeting, for example, Chris Voigt from

the University of California, San Francisco, reported that his lab has engineered *Salmonella* bacteria to make and secrete protein components — a difficult task made easier by the lab's deliberate simplification of the genetic code for the proteins. And Jeff Boeke from the Johns Hopkins University in Baltimore, Maryland, has built, from scratch, a retrotransposon — a

genetic element capable of jumping around the genome. Boeke's retrotransposon is designed to be a far more effective 'jumper' than natural retrotransposons, and indeed, inserts itself into many more places in the genome.

Laboratory life

Such early success has synthetic biologists dreaming of far grander projects. Boeke, for instance, is gathering support for the idea of engineering synthetic yeast organisms, making them compete against each other, and watching which strains evolve. George Church's lab at the Massachusetts Institute of Technology (MIT) in Boston, is making synthetic stretches of *Escherichia coli* DNA with the goal of 'optimizing' the bacterium's genome and reconstituting it from scratch. Drew Endy at MIT has already done this with a T7 bacteriophage. And while other scientists have built whole viruses from synthetic DNA by simply ordering viral DNA from a gene-synthesis company (see *Nature* **418**, 265; 2002), synthetic biologists envisage something different. Their goal is to strip genomes down to their essential parts to try to learn more



CONFERENCE REPORTS

More news and diary entries from this week's Synthetic Biology 2.0 meeting.

www.nature.com/news

about the principles behind the architecture of life.

"The time is right for this," Boeke says. "This would allow us to answer functional questions we can't ask in any other way."

But as exciting as these ideas are, they also raise concerns about the proper use of such technology. A coalition of organizations traditionally concerned about biotechnology, including the ETC Group, Greenpeace and GeneWatch UK, submitted an open letter to the synthetic biologists and to media before the meeting, protesting against the biologists' self-regulating approach. "We believe that this potentially powerful technology is being developed without proper societal debate concerning socioeconomic, security, health, environmental and human rights implications," the letter states.

Enemy agents?

Companies have already run up against biosecurity issues. Jeremy Minshull, president and co-founder of the company DNA 2.0, and Hans Bügl of Geneart, based in Regensburg, Germany, have both turned away orders for potentially hazardous DNA. In Geneart's case, an Indian customer ordered a stretch of DNA that is banned for export by the German government. In Minshull's case, a client ordered a pathogen on the 'select agent' list of organisms issued by the US Centers for Disease Control and Prevention (CDC). No one at the agency could tell Minshull how to handle this sensitive order, so he cancelled it. In both cases, the companies believed the customers had good intentions. The DNA 2.0 case is particularly worrying, says Gerald Epstein of the Center for Strategic and International Studies in Washington DC: "CDC implements a select agent list, but it doesn't say anything about what happens if somebody suspects somebody else of doing something bad."

Ultimately, the synthetic biologists said at the meeting, the best way to deal with these issues is to pursue projects that benefit society. For instance, Jonathan Eisen of the University of California, Davis, said the community can make a positive contribution to biodefence by engineering ways to tag and track DNA. This might help pinpoint the source of biological attacks. Such efforts could win public support for the technology, and this will be crucial for the field's future, said David Baltimore, president of the California Institute of Technology in Pasadena: "We have to remind everyone that we are working towards a greater good."

Erika Check

Mars explorers seek spot for touchdown

More than 125 planetary scientists will gather in Pasadena, California, next week to begin planning NASA's next steps on Mars — or rather, where that step should fall. The workshop will rank more than 40 candidate landing sites for the Mars Science Laboratory (MSL), a rover slated to depart for the planet in September 2009. There will be no hasty decisions: NASA plans three more workshops after this one, and will choose the final site a month before launch.

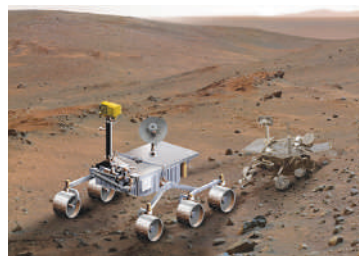
NASA says the MSL will improve in every way on the Spirit and Opportunity rovers now exploring Mars. Its ten instruments weigh a total of 75 kg, compared with five instruments weighing 9 kg on the current rovers. It will be able to land within a 20-km-diameter target circle, anywhere within 60° north or south of the martian equator, instead of being restricted to a narrow band around the planet's middle. And it will be able to negotiate rougher and steeper terrain than Spirit and Opportunity and travel further — at least 20 kilometres.

With so many options for where to land, workshop co-chairman John Grant of the Smithsonian National Air and Space Museum's Center for Earth and Planetary Studies feared scientists would suggest thousands of candidate sites. "We were

quite relieved to find out that the number of sites is manageable," he says. The first workshop's goal is to rank the sites as high, medium or low priority for detailed study using cameras and other instruments orbiting Mars.

The MSL will continue the work of its predecessors in tracing the history of water on Mars. Geologists would prefer to explore an area where many layers of rock are exposed, to give a large cross-section

through time. The less dust covering the rocks, the better. The site must also be navigable for a wheeled rover, without too many obstacles. And particularly prized will be areas where scientists believe water once settled in quiet pools that may have provided habitats for ancient martian life (see



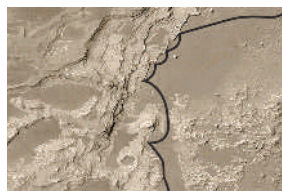
Top dog: the Mars Science Laboratory (left) will dwarf today's rovers (right).

'Candidate landing sites').

Another improvement on past rover missions is that scientists will know more about the landing sites in advance. Since Spirit and Opportunity arrived on Mars in January 2004, Mars Express and the Mars Reconnaissance Orbiter (MRO) have begun their own study of the surface. The MRO's High Resolution Imaging Science Experiment may be able to spot outcrops from orbit of the kind that Spirit and Opportunity have visited.

Tony Reichhardt

CANDIDATE LANDING SITES



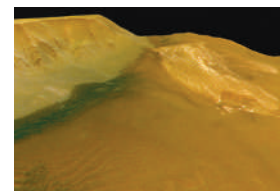
SEDIMENTARY FANS

A rover in the Eberswalde crater could roam (black line) among fan-shaped sedimentary deposits at the end of flow channels, pausing at rock exposures. Eberswalde's features suggest that there was once standing water there. Plus, the crater floor is flat, making for easy travel.



CRATER FLOORS

At least 50 metres of water is believed to have pooled in some parts of Holden crater, one of a string of large, ancient craters in Mars's southern hemisphere. Fan-shaped sediments near the crater's centre suggest calm waters where life might once have taken hold.



SULPHATE DEPOSITS

The OMEGA spectrometer on the European Mars Express has found minerals such as sulphate and phyllosilicates that form in water, making them candidates for further exploration. The sulphate deposits pictured here are in a canyon called the Juventae Chasma.

T.R.

Reconstructed skeleton of an 85 million-year-old plesiosaur, the fossilized bones of which were found near Tokyo in 1968.



Teenager waits 40 years for recognition

TOKYO

Tadashi Suzuki was a fossil-mad high-school student when he discovered some ancient bones in a riverbank near his home. He wrote to a local palaeontologist about them, thinking he might have discovered a plesiosaur.

That was in 1968. Almost four decades later, his find has finally been confirmed as a new species. Although the time taken in this case is extreme, experts say it highlights the problems many palaeontologists face in raising the funds to characterize and classify new discoveries.

Suzuki, now 54, says he has been fascinated by fossils since his childhood. When he was 17, he read about a fossil plesiosaur — a long-necked reptile that flourished in the world's oceans during the Cretaceous era — being found in his home town of Iwaki City, about 200 kilometres north of Tokyo. "I thought I could find another," he says. He went down to the same riverbank and discovered more bones.

Rather than disturb the fossils, he wrote to Ikuro Obata, a palaeontologist in Tokyo, to tell him about the find. Obata's colleague Yoshikazu Kasegawa took on the project.

It took about seven years to excavate the fossil bones, clean them up and assemble them into a three-metre-long plesiosaur skeleton,

which is around 70% complete. Hasegawa believed Suzuki's find to be a new species, but research then slowed almost to a halt.

Proving that a specimen is a new species requires comparing it to all other known specimens and showing it is sufficiently different. Hasegawa, now head of the Gunma Museum of Natural History, says he couldn't afford all the relevant references as well as travel to foreign museums and libraries. The project finally took off in 2003, when Tamaki Sato joined the team. A researcher at the National Science Museum in Tokyo, she had studied plesiosaurs in Canada and had data from other specimens.

The paper describing Suzuki's plesiosaur as a new species was finally published this month (T. Sato, Y. Hasegawa and M. Manabe *Palaeontology* 49, 467–484; 2006). "I am delighted to see the research published," says Suzuki, now chief researcher at his local ammonite centre. "I think the 38 years were inevitable."

The team has called it *Futabasaurus suzukii* and reports that it is an elasmosaur, the main type of plesiosaur. Elasmosaurs lived all over the world, but few specimens from Asia have been

well preserved. The Japanese find is 85 million years old, the oldest specimen in the northern Pacific confirmed to be an elasmosaur.

Plesiosaurs have been little studied compared with other species groups that were around at the time, such as dinosaurs and invertebrates. Many details, such as how they swam, remain a mystery (see 'Mystery and myth behind the plesiosaur'). "It's a nice addition to the story of plesiosaurs," says Leslie Noé, a specialist in ancient reptiles at the University of Cambridge's Sedgwick Museum in Britain. "New plesiosaurs give us a unique insight into the biodiversity of the Cretaceous seas."

"Large vertebrate fossils are seen as high profile, but this rarely results in money for study and display."

Noé says it would take about five years to get a fossil find classified if researchers had sufficient time, money and personnel. But the situation faced by Hasegawa and his colleagues will sound familiar to many palaeontologists, who find it increasingly difficult to get funds, he adds. "Large fossil vertebrates are seen as high profile by the public, but this rarely results in the money required to study and display the specimens becoming available."

Ichiko Fuyuno

NATIONAL SCIENCE MUSEUM, TOKYO

Mystery and myth behind the plesiosaur

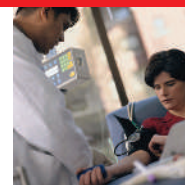
When the ancestors of plesiosaurs abandoned land for water more than 215 million years ago, they gave their descendants an unusual mechanism of propulsion. Almost all other known water-dwellers use a tail to move, but these reptiles swam by flapping or rotating flipper-like appendages that

evolved from what were previously limbs. The exact mechanism is still the subject of debate.

Larger plesiosaurs grew to more than 10 metres in length, making them the biggest aquatic animals of the time. Their long, flexible necks and strong jaws are thought to have been used to pluck fish

from passing shoals. But that picture is being reconsidered following the publication last year of a plesiosaur fossil that had remains of bottom-dwelling creatures, such as crustaceans, in its stomach (see C. R. McHenry, A. G. Cook and S. Wroe *Science* 310, 75; 2005).

The reptiles are also a candidate for unfounded but popular theories that their descendants survive to this day. Scientists dismissed claims that a Japanese trawler had caught a decaying plesiosaur in the late 1970s and the Scottish press sometimes links plesiosaurs with the Loch Ness monster. **Jim Giles**


**PHARMA FIRMS TOLD TO
END SECRECY IN TRIALS**

Early stages of drug testing should be made public, say public-health experts.

www.nature.com/news

Election fever inflames the US stem-cell debate

In the run-up to autumn's Senate elections, stem-cell research has become something of a political football. This is raising the profile of bills that could loosen restrictions on federal funding for the research, but some fear that scientific arguments are being overshadowed by politicians' posturing.

In November, voters will decide who occupies 33 seats in the Senate and 435 in Congress, and stem cells are poised to play a decisive role in some key races.

Last May, the US House of Representatives voted overwhelmingly to loosen restrictions on federal funding of human embryonic stem-cell research. Since then, the Republican Senate majority leader, Bill Frist, has faced demands from stem-cell research advocates to deliver on a promise to bring the issue to the Senate floor (see *Nature* **436**, 608; 2005).

Observers say Frist must factor numerous political considerations into his decision about when and how to bring the issue before the Sen-

ate, including how it would be viewed by the White House and how it would affect his own career. The issue has become difficult for the Republican party because although a majority of voters support the work, it is viewed as immoral by the party's conservative core.

Washington is buzzing with rumours. Last week, Associated Press reported that senators are deciding whether to vote on a package of three stem-cell bills later this summer. But no senator has confirmed the story, leaving advocates uncertain what this will mean for research. "Only the details will determine whether it is good or bad for researchers and, ultimately, patients," says Kevin Wilson, director of public policy at the American Society for Cell Biology.

The package would comprise the bill passed by the House and two others, one of which was introduced on 5 May by Pennsylvania senators Arlen Specter and Rick Santorum, both Republicans. Their bill supports research that tries to find sources for human embry-

onic stem cells that do not harm embryos.

The deal might help Santorum, who faces a bruising re-election battle this autumn. But it is unclear how the bill would affect other races.

The White House, meanwhile, stands by President George W. Bush's threat to veto any bill that loosens restrictions on federal funding for stem-cell research. Some say such a veto may actually help Bush win some badly needed points. The Christian right is beating him up, says one research advocate, speaking off the record. "The White House needs to put a stop to that, and this would do it."

Such a scenario would not help researchers much. And some in the research community are nervous about the effect this political manoeuvring could have on science. "These are complicated, cutting-edge scientific issues, and they need to be considered in a thoughtful and educated manner," says Wilson. "In general, this is not the best way to legislate."

Erika Check

SPECIAL REPORT

Named and shamed

As accusations of scientific misconduct in China become rife, some fear persecution reminiscent of that used in the Cultural Revolution.

Chinese science risks being sliced up by a double-edged sword: rampant scientific misconduct on the one hand, and persecution based on false accusations on the other.

The lack of confidence in official mechanisms for properly investigating fraud has led to increased reliance on websites that challenge the records and publications of Chinese scientists. But many are concerned about the damage such untested allegations can cause; more than 100 Chinese scientists based in the United States have sent an open letter to the Chinese government, asking it to set up mechanisms to ensure that claims of scientific misconduct are investigated fairly.

China admits it faces a serious problem with scientific misconduct, including plagiarism, and the fabrication and falsification of data. The scale of the problem is unknown, but a recent spate of allegations has drawn attention to the issue.

In March, Hui Liu, the vice-dean of Tsinghua University medical school in Beijing, was fired, following claims that he had boosted his publication list with papers by another H. Liu (see *Nature* 440, 728; 2006). Liu has reportedly denied the charges and blamed the mix-up on a clerical error. In April, Sichuan University in Chengdu was criticized by the Chinese media for finding one of its professors innocent of fabricating a paper; the paper has been under

attack since its publication in 2000. And two weeks ago, Jin Chen of Shanghai's Xi'an Jiaotong University, whose announcements of one of China's first digital signal-processing chips in 2003 stoked patriotic fervour, was condemned by his university for faking research and stealing designs from a foreign company.

In all three cases, a popular Chinese-language website known as New Threads (www.xys.org), which has a reputation for disclosing scientific fraud in China, played a key role in fuelling public outcry.

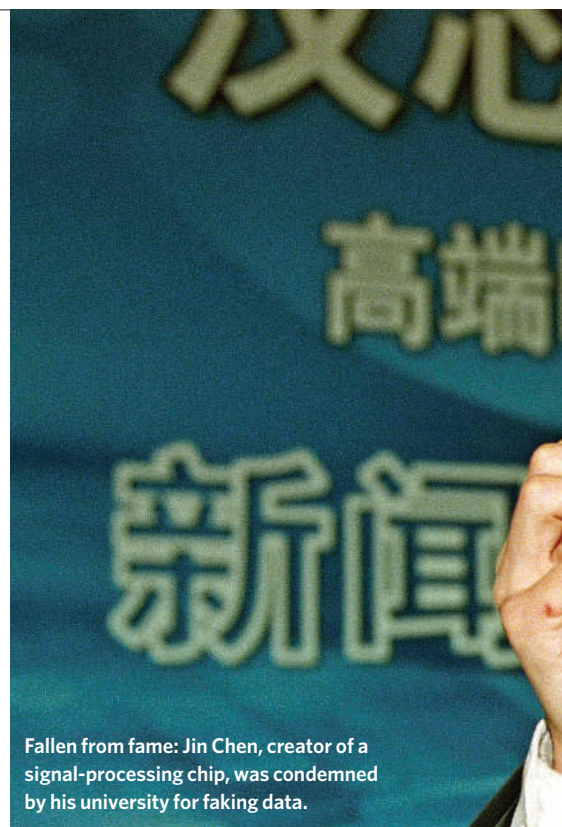
In the first two cases, postings of the accusations on New Threads led to the Chinese media picking up on the stories. And the website's owner, Shi-min Fang, a biochemist based in San Diego, California, claims he was the first to post the name of Chen's company which supposedly re-labelled foreign chips.

The power of the website to implicate scientists in the absence of adequate formal mechanisms of investigation has put it at the centre of concerns over claims of misconduct.

Xin-Yuan Fu, an immunologist at Indiana University in Indianapolis, says it was the Sichuan University case that drove him to write a letter to key science-policy officials, including China's science and technology minister and the head of the Chinese Academy of Sciences, asking them to take action. The letter struck a chord among his peers — within five days of

"The recent self-investigation into alleged fraud at Sichuan University is a total joke."

AP



Fallen from fame: Jin Chen, creator of a signal-processing chip, was condemned by his university for faking data.

circulating it to other Chinese biologists based in the United States, Fu's letter had collected 120 signatures, including those of two researchers in China. "I was overwhelmed," says Fu.

After noting the need to expose all types of misconduct, the letter focuses on the problem of unfounded allegations, particularly those that attack scientific claims without giving evidence of faulty laboratory procedures. It ends by condemning the tendency to make "personal attacks anonymously in public... in the absence of proper investigation".

Fu says the Sichuan University incident is a case in point. Yuquan Wei, vice-president of the university, published a paper in *Nature Medicine* in 2000 detailing the use of foreign endothelial cells as a vaccine to prevent tumour growth. The paper claimed success in mice and suggested the technique could work in humans (*Nature Med.* 6, 1160–1166; 2000).

But Lusheng Si, an immunopathologist at Xi'an Jiaotong University who first came across the paper when reviewing a grant proposal by Wei in 2001, suspected that it contained fabricated data. On 26 March this year, after hearing that Wei was using the paper to request a further large grant, Si attacked the paper on New Threads.

The letter led to a media fury in China and an investigation by Wei's university. Sichuan concluded that Wei had committed no offence, and that the dispute over Wei's



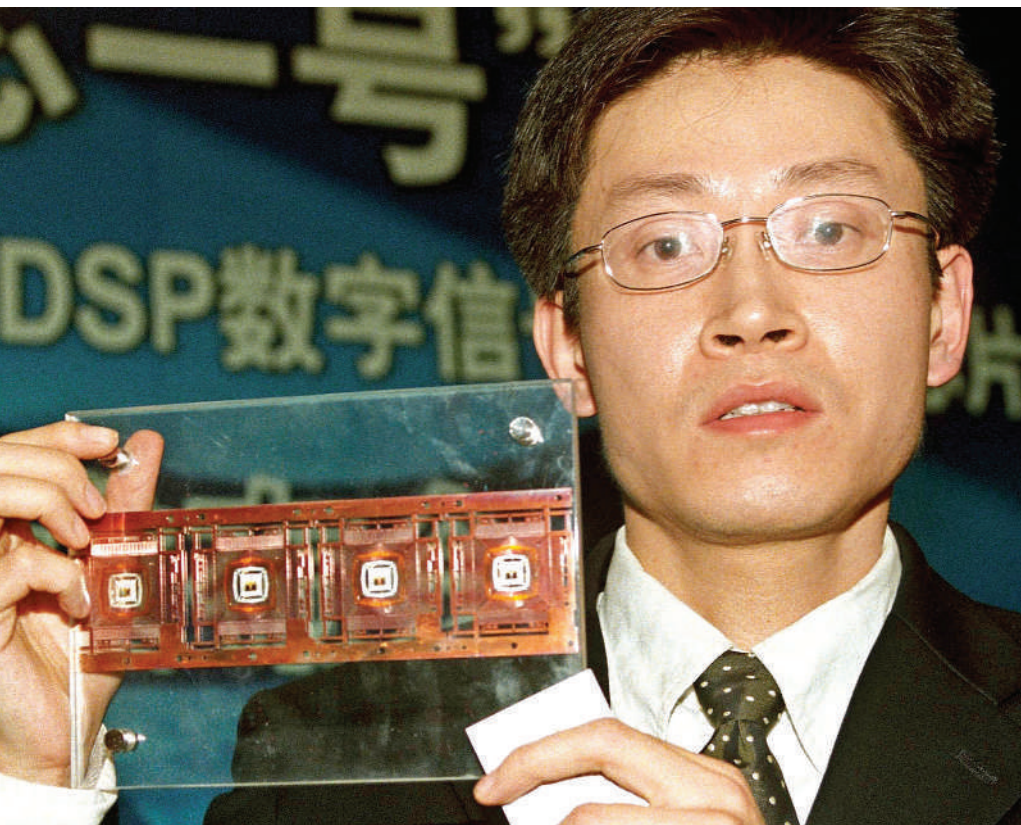
Accusations of scientific fraud posted on websites remind some of the posters used to persecute 'government enemies' in the 1970s.

J. ANDANSON/SYGMA/CORBIS

**TSUNAMI TEST RUN**

Read our interview with Brian Yanagi, who oversaw this month's test of the Pacific tsunami warning system.

www.nature.com/news



research was simply a run-of-the-mill academic disagreement. The media in China has continued to criticize Wei and Sichuan University, but many scientists think Si's attack was irresponsible and based on unsound interpretation of scientific concepts and procedures.

Si contends, for example, that the mouse immune system should respond to all proteins in foreign cells, whereas Wei's paper suggests that immunized mice selectively respond to a few antigens. "This violates a fundamental law of immunology," Si says.

But Lieping Chen, an immunologist at Johns Hopkins University School of Medicine in Baltimore, Maryland, and a signatory to Fu's letter, disagrees with Si. Chen says that a selective immune response to one or a few foreign proteins is an aspect of well-known phenomenon known as immunodominance.

Si also questions the number of mice Wei used, estimating this to be around 40,000. "This is too big to believe," he says. Wei, backed by Chen, says Si has miscalculated the number, and that less than 5,000 mice were actually used.

But even those who defend Wei admit that his response hasn't helped. For example, Si claims that Wei has so far refused to release his raw data, which most agree would settle the issue. Wei told *Nature*, "I did not say I cannot release raw data for inspection", but he has not clarified whether he will make his

data available. He has denied all misconduct.

The university's investigation into the matter has failed to convince many that the truth won out, mainly because it lacked transparency. "The recent self-investigation into alleged fraud at Sichuan University is a total joke," says Mu-ming Poo, a neurobiologist at the University of California, Berkeley, and head of the Institute of Neurosciences in Shanghai. *Nature's* request for details on the university procedure and an introduction to members of the investigation committee was referred to Wei; as *Nature* went to press he had not provided any information about the investigation.

Poo believes the incident is indicative of the fact that most Chinese universities lack the capacity to investigate one of their own. "The outcome is likely to be influenced by the university's own interests, such as protecting its reputation," he says.

Fu's letter, sent on 8 May, calls for greater involvement of higher-level funding bodies such as the science ministry, the Chinese Academy of Sciences (CAS) and the Natural Science Foundation of China (NSFC).

These institutions already have investigatory bodies. The CAS established its ethics committee in 1997 and drafted guidelines in 2001. The NSFC committee, established in 1998, says it investigated 445 allegations of

misconduct in its first five years (out of an estimated 30,000 projects that it funded during that time). In the most severe cases, the committee indefinitely blocks perpetrators from applying for funds.

But many scientists feel these committees are ineffective, and a lack of confidence in their ability to settle matters is driving those with grievances to publish them on the Internet. For example, Si says he considered sending his complaint to the CAS or to the science ministry, but he was unable to find contact details for either. So he posted his accusation on New Threads instead. *Nature's* attempts to contact the committees of the CAS and the NSFC were also unsuccessful.

"It is the [effective] absence of such formal mechanisms that makes New Threads important," says Fu. But Fu, a human-rights advocate, is worried that the media frenzy following irresponsible web-based accusations, particularly by those who don't identify themselves, harkens back to China's 'big letter' posters or 'dazibao'.

These wall-mounted handwritten posters were used to persecute those considered enemies of the government during the Cultural Revolution in the 1970s. "Anyone could write anything, and people would read it and assume it was right," says Chen. "It would be a terrible thing to go through again, in academia."

Fang, who has been widely praised since setting up his website in 2001 for exposing bad science and trying to raise the profile of research ethics in China, defends his postings. He says he only accepts about 10% of submitted letters, and that he only publishes allegations from correspondents who identify themselves to him. He adds that he does some preliminary

investigation and sometimes asks outside experts for their opinions.

But several scientists have written to *Nature* to express concern over how powerful Fang's website has become, saying they are afraid to be named for fear of becoming his enemy.

Ideally, Fu says he would like to see China establish a new agency staffed by experts trained in scientific misconduct that could investigate claims of fraud, akin to the US Office of Research Integrity. That would certainly be necessary to resolve the case of Si versus Wei, says *Nature Medicine's* editor-in-chief Juan-Carlos Lopez. "There's been enough of this 'he said, she said' nonsense," says Lopez. "It's time for the competent authorities to get involved."

How likely that is to happen is unclear. Fu and his co-signatories have yet to receive any response from the Chinese authorities. ■

David Cyranoski

"There's been enough of this 'he said, she said' nonsense."

ON THE RECORD

“It would take somebody with real balls to say to Einstein, ‘Look, this is wrong.’”

A physicist praises Ernst Gabor Straus — a young colleague of Albert Einstein's, whose collection of correspondence with the legend will go on sale next month.

“I saw this ad that said ‘Get paid to stay in bed.’ I thought it was a scam.”

Erin Peterson tells how she came to spend three months lying down for a NASA bedrest study.

Sources: *Guardian*, *Seed*

SCORECARD

Spatial awareness
It turns out that NASA's DART mission failed last year because it accidentally accelerated into the satellite it was meant to be circling.

Whale songs
Anti-whaling activists launch a competition for the best music made from the remixed sound of singing whales.

Nazi researchers
A German scientist is ousted from the International Space Hall of Fame after revelations that he did experiments on prisoners in Dachau concentration camp.

Public image
The US National Science Foundation hires a public-affairs chief — former communications director for vice-president Dan Quayle. In 1991, Quayle was awarded the satirical Ig Nobel prize for “demonstrating, better than anyone else, the need for science education”.

NUMBER CRUNCH

Would you rather be fat or alive?
A survey of 4,300 people says...

25% of people would rather be unable to have children than obese.

30% would rather be divorced.

46% would rather give up a year of their life than be obese.

Source: Schwartz, M. B., Vartanian, L. R., Nosek, B. A. & Brownell, K. D. *Obesity* 14, 440–447; (2006).

ITER



Grand plan: a prototype reactor is poised to be built at Cadarache, but is there time to alter the design?

Physicists plead to make final tweak to fusion experiment

Just as scientists and engineers from seven nations put the finishing touches to their design for a multi-billion-dollar experiment, an idea comes along that could improve it. The dilemma facing those working on the international fusion reactor, known as ITER, is whether or not to incorporate the change.

ITER will cost roughly US\$6 billion to build — and aims to be the first fusion reactor to produce more power than it consumes. It will work by using a web of carefully constructed magnetic fields to suspend a plasma gas, which is heated to hundreds of millions of degrees.

A tweak to those fields is now being proposed, based on findings by plasma physicist Todd Evans and his team at General Atomics in San Diego, California. Evans's group successfully demonstrated a technique for preventing dangerous, lightning-like plasma discharges that could damage key parts of the reactor. These are caused when plasma builds up in the weakest areas of the magnetic field that contains it. “Think of squeezing a balloon of water — the balloon bulges out through your fingers,” says Evans. By modifying a reactor machine at General Atomics, called the DIII-D tokamak, the team made the rigid magnetic fields a little fuzzy — allowing excess plasma to gently leak away rather than bursting out. “In a sense, it's a beautiful concept,” Evans says.

The idea, which has been published in *Nature Physics* (T. E. Evans *et al.* *Nature Phys.* doi:10.1038/nphys312; 2006), could help ITER to succeed more quickly. But it comes at a cost: the technique would probably require expensive superconducting coils to be placed near or

inside the containment vessel, where space is limited and punishing radiation wears out equipment quickly.

ITER researchers are now mulling over how to work such a breakthrough into the design, which originated more than 20 years ago and has already been through several iterations. The change will have to wind its way through several review committees before receiving final sign-off by designers in the countries funding it: the United States, the European Union, Japan, the Russian Federation, India, China and South Korea. It will also compete with other ideas in the field that could help prevent violent disruptions in the machine or extend the time it can hold fusing plasmas.

Trying to get scientists and engineers to decide which changes to include in the final design and which to save for ‘upgrades’ isn't going to be easy, says Ned Sauthoff, a physicist at Princeton Plasma Physics Laboratory in New Jersey and head of ITER's US team.

“ITER is not just a physics experiment, it's also an experiment in large-scale project management.”

In the meantime supporters of the new concept will continue to conduct experiments, says Philippe Ghendrih, a theoretical physicist at the French Atomic Energy Commission in Cadarache, where further work is being done to integrate the idea into existing ITER designs. “If several other tokamaks can demonstrate the effect, then it will be easier to convince the engineers.”

A final design review will probably begin in the autumn, and should take about a year. ■

Geoff Brumfiel

“In a sense, it's a beautiful concept.”

Congress wants more cash for health, less for nuclear

US biomedical research may emerge as a winner from the US Congress's budget machinations — or at least not as much of a loser as had been expected.

On 18 May, the House of Representatives called for an extra \$7 billion to be given above the amount requested by President George W. Bush in February to the 2007 budgets for labour, health and education. That could include an unspecified amount for the National Institutes of Health, which faces its first budget cut in more than three decades. The Senate has also called for the boost, but the numbers are not binding and the exact amount will be decided in committees over the summer.

Bush's American Competitiveness Initiative seems likely to get the money he requested, including increased funding for physical sciences at the Department of Energy (see *Nature* 439, 644–645; 2006). But the president's plans for nuclear science are faring less well. On 18 May, the House appropriated just \$120 million — \$130 million less than requested — for the administration's fuel-leasing and reprocessing scheme, the Global Nuclear Energy Partnership.

Study heightens fears of second wave of vCJD

A study of human tissue samples has increased speculation that there could be another wave of prion-disease cases.

In Britain, the number of cases of variant Creutzfeldt–Jakob disease (vCJD), which is contracted by eating infected beef, has been falling since 2000, when 28 people died. All known cases have occurred in patients carrying a genetic variation known as *MM*. But the study (J. W. Ironside *et al.* *BMJ* 332, 1186–1188; 2006) has identified a second

group of susceptible people — those with a variation known as *VV*.

Examination of more than 12,000 tissue samples removed in British hospitals between 1995 and 2000 revealed two patients with the *VV* variation who were infected with the malformed proteins, called prions, that cause vCJD.

The result shows that *VV* carriers can become infected, and raises the possibility that they simply incubate the disease for longer. In an editorial accompanying the paper, public-health experts say the results demonstrate the urgent need for a test for vCJD, so infected people can be excluded from donating blood or other organs.

NIH taken to court over proposed biodefence lab

Several environmental and civil-rights groups have sued the US National Institutes of Health (NIH) to block the construction of a biodefence laboratory in Boston. They allege that the \$178-million facility, for research on deadly pathogens, was improperly located in a poor community.

Local activists have opposed the Boston University lab for years (see *Nature* 428, 785; 2004). But in February, officials from the NIH and the federal government approved the construction of the laboratory, one of a series of biodefence labs being built nationwide.

The lawsuit argues that the agency did not conduct proper environmental reviews and dismissed the risks to public health. Lawyers say the government is required to exercise special caution when reviewing such facilities in minority communities, which often bear the brunt of environmental problems. NIH officials declined to comment on the pending litigation.

Italy appoints chemist in bid to embrace science

One of Italy's most-cited scientists has been made a minister in the government of the newly elected prime minister, Romano Prodi.

Luigi Nicolais, a polymer chemist at the University of Naples 'Federico II', will head the ministry for public function and innovation, which is tasked with modernizing the machinery of government. Nicolais is a member of the 2003 Group, which comprises the 63 most-cited Italian scientists. The group was formed three years ago to promote the interests of research.

Mathematician Luciano Modica, former president of the Conference of Italian University Rectors, has also been given a government job, as undersecretary to the



On the rise again? A second group of people could get variant Creutzfeldt–Jakob from infected beef.

PA/EMPHICS

Japan offers universities and business a free launch

JAXA

Japan's space agency will begin offering free rides for homemade satellites aboard its H-IIA rocket (pictured).

Universities and private companies that are developing satellites for research purposes can now apply for a free launch from the Japan Aerospace Exploration Agency.

The agency is planning to launch a few such satellites at least once a year, with the first one scheduled for 2008. It

hopes the programme will help beef up the country's satellite expertise.

The move was praised by space-exploration enthusiasts in Japan, some of whom have travelled to other countries to launch small satellites at their own expense.



minister of universities and research. The new government claims it is committed to building up the research base in Italy, support for which had been reduced under Silvio Berlusconi.

India maps out plans for navigational satellites

India has announced plans to launch its own constellation of navigational satellites. The move is meant to end the country's reliance on the Global Positioning System (GPS) satellites, which are run by the US defence department.

India's cabinet sanctioned the \$310-million project on 9 May. The system would consist of seven satellites covering the Indian subcontinent, to be built and launched starting in 2008. The Indian Space Research Organisation hopes to have the satellites in place by 2011.

Officials say the project is independent of India's negotiations to join two other navigation systems — Europe's fledgling Galileo project and Russia's half-completed GLONASS system.

Chile wins race to host sky-survey telescope

A powerful all-sky survey telescope proposed to come online in 2012 will be located in Chile, not at a competing site in Mexico, project officials announced last week.

A site selection committee for the 8.4-metre Large Synoptic Survey Telescope (LSST) chose Cerro Pachón in the Andes, in part because it has more existing support facilities for astronomy than the rival San

Pedro Mártir site in Baja California (see *Nature* 439, 377; 2006). Cerro Pachón is already the home of the Gemini South and Southern Astrophysical Research telescopes.

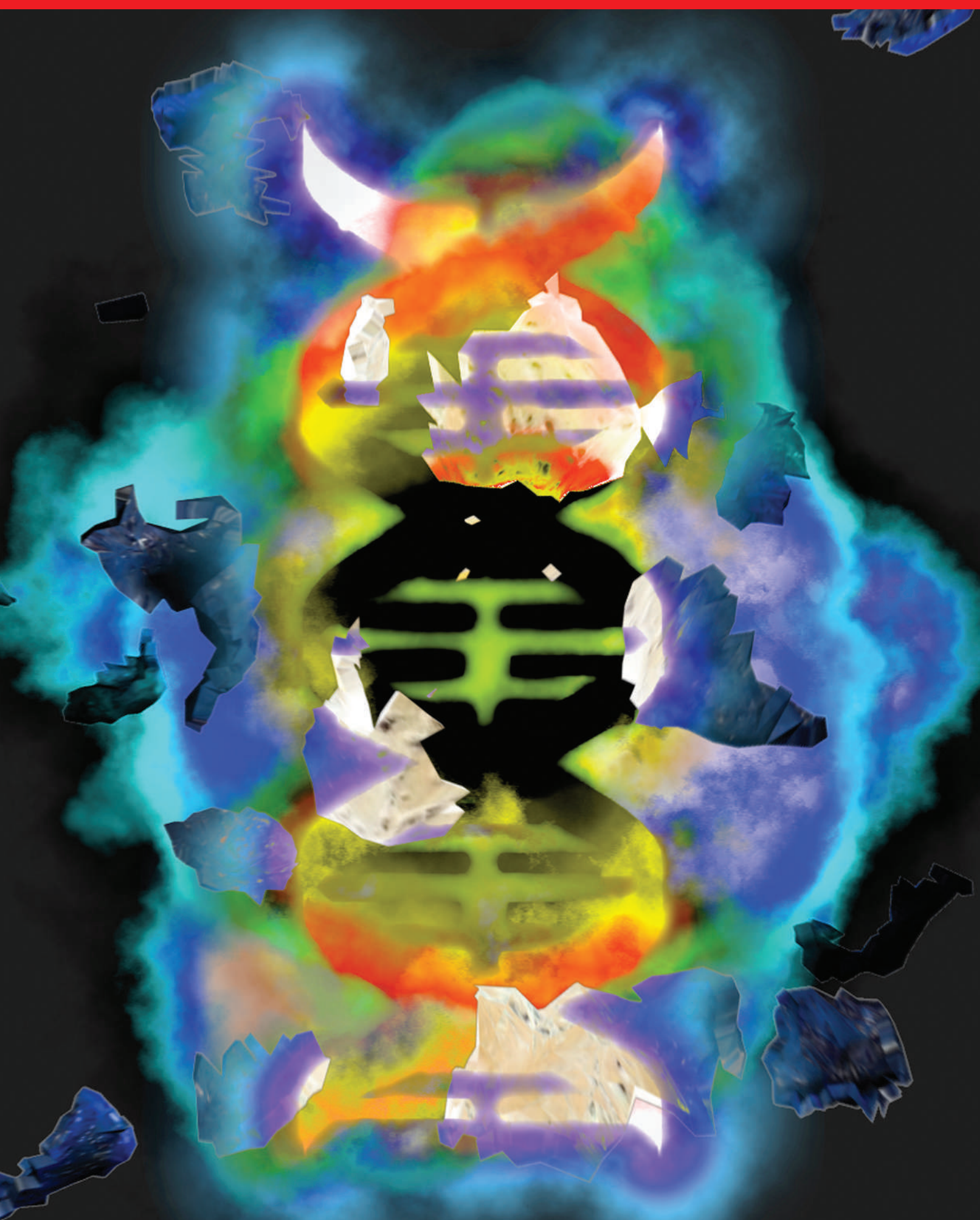
The LSST will be able to survey the entire visible sky every three days, building up an unprecedented catalogue of high-resolution images that could be used to study everything from dark matter to Earth-approaching asteroids. The US government has yet to commit funds for the \$270-million telescope, however. Project backers plan to submit a construction proposal to the National Science Foundation by the end of this year.

Sussex saves threatened chemistry department

At least one UK chemistry department received some good news last week. The University of Sussex announced it has reversed earlier plans to close the department and replace it with a chemical-biology department.

The university's ruling bodies approved the deal on 15 May. It will create a merged department of chemistry and biochemistry, saving the jobs of the current staff and allowing Sussex to continue offering chemistry degrees to undergraduates. But the university will not, at least at first, replace staff lost over the past few years.

Reports of the department's proposed closure triggered an unprecedented fundraising drive this spring to save it (see *Nature* 441, 12–13; 2006). Other departments across the country have also been suffering because of a fall in the number of students enrolling in chemistry.



WHAT IS A GENE?

The idea of genes as beads on a DNA string is fast fading. Protein-coding sequences have no clear beginning or end and RNA is a key part of the information package, reports **Helen Pearson**.

'Gene' is not a typical four-letter word. It is not offensive. It is never bleeped out of TV shows. And where the meaning of most four-letter words is all too clear, that of gene is not. The more expert scientists become in molecular genetics, the less easy it is to be sure about what, if anything, a gene actually is.

Rick Young, a geneticist at the Whitehead Institute in Cambridge, Massachusetts, says that when he first started teaching as a young professor two decades ago, it took him about two hours to teach fresh-faced undergraduates what a gene was and the nuts and bolts of how it worked. Today, he and his colleagues need three months of lectures to convey the concept of the gene, and that's not because the students are any less bright. "It takes a whole semester to teach this stuff to talented graduates," Young says. "It used to be we could give a one-off definition and now it's much more complicated."

In classical genetics, a gene was an abstract concept — a unit of inheritance that ferried a characteristic from parent to child. As biochemistry came into its own, those characteristics were associated with enzymes or proteins, one for each gene. And with the advent of molecular biology, genes became real, physical things — sequences of DNA which when converted into strands of so-called messenger RNA could be used as the basis for building their associated protein piece by piece. The great coiled DNA molecules of the chromosomes were seen as long strings on which gene sequences sat like discrete beads.

This picture is still the working model for many scientists. But those at the forefront of genetic research see it as increasingly old-fashioned — a crude approximation that, at best, hides fascinating new complexities and, at worst, blinds its users to useful new paths of enquiry.

Information, it seems, is parceled out along chromosomes in a much more complex way than was originally supposed. RNA molecules are not just passive conduits through which the gene's message flows into the world but active regulators of cellular processes. In some cases, RNA may even pass information across generations — normally the sole preserve of DNA.

An eye-opening study last year raised the possibility that plants sometimes rewrite their DNA on the basis of RNA messages inherited from generations past¹. A study on page 469 of this issue suggests that a comparable phenomenon might occur in mice, and by implication in other mammals². If this type of phenomenon is indeed widespread, it "would have huge implications," says evolutionary geneticist

Laurence Hurst at the University of Bath, UK.

"All of that information seriously challenges our conventional definition of a gene," says molecular biologist Bing Ren at the University of California, San Diego. And the information challenge is about to get even tougher. Later this year, a glut of data will be released from the international Encyclopedia of DNA Elements (ENCODE) project. The pilot phase of ENCODE involves scrutinizing roughly 1% of the human genome in unprecedented detail; the aim is to find all the sequences that serve a useful purpose and explain what that purpose is. "When we started the ENCODE project I had a different view of what a gene was," says contributing researcher Roderic Guigo at the Center for Genomic Regulation in Barcelona. "The degree of complexity we've seen was not anticipated."

Under fire

The first of the complexities to challenge molecular biology's paradigm of a single DNA sequence encoding a single protein was alternative splicing, discovered in viruses in 1977 (see 'Hard to track', overleaf). Most of the DNA sequences describing proteins in humans have a modular arrangement in which exons, which carry the instructions for making proteins, are interspersed with non-coding introns. In alternative splicing, the cell snips out introns and sews together the exons in various different orders, creating messages that can code for different proteins. Over the years geneticists have also documented overlapping genes, genes within genes and countless other weird arrangements (see 'Muddling over genes', overleaf).

Alternative splicing, however, did not in itself require a drastic reappraisal of the notion of a gene; it just showed that some DNA sequences could describe more than one protein. Today's assault on the gene concept is more far reaching, fuelled largely by studies that show the pre-

viously unimagined scope of RNA.

The one gene, one protein idea is coming under particular assault from researchers who are comprehensively extracting and analysing the RNA messages, or transcripts, manufactured by genomes, including the human and mouse genome. Researchers led by Thomas Gingeras at the company Affymetrix in Santa Clara, California, for example, recently studied all the transcripts from ten chromosomes across eight human cell lines and worked out

precisely where on the chromosomes each of the transcripts came from³.

The picture these studies paint is one of mind-boggling complexity. Instead of discrete genes dutifully mass-producing

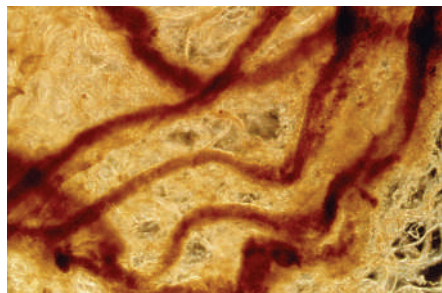
identical RNA transcripts, a teeming mass of transcription converts many segments of the genome into multiple RNA ribbons of differing lengths. These ribbons can be generated from both strands of DNA, rather than from just one as was conventionally thought. Some of these transcripts come from regions of DNA previously identified as holding protein-coding genes. But many do not. "It's somewhat revolutionary," says Gingeras's colleague Phillip Kapranov. "We've come to the realization that the genome is full of overlapping transcripts."

Other studies, one by Guigo's team⁴, and one by geneticist Rotem Sorek⁵, now at Tel Aviv University, Israel, and his colleagues, have hinted at the reasons behind the mass of transcription. The two teams investigated occasional reports that transcription can start at a DNA sequence associated with one protein and run straight through into the gene for a completely different protein, producing a fused transcript. By delving into databases of human RNA transcripts, Guigo's team estimate that 4–5% of the DNA in regions conventionally recognized as genes is transcribed in this way. Producing fused transcripts could be one way for a cell to generate a greater variety of proteins from a limited number of exons, the researchers say.

Many scientists are now starting to think that the descriptions of proteins encoded in DNA know no borders — that each sequence reaches into the next and beyond. This idea will be one of the central points to emerge from the ENCODE project when its results are published later this year.

Kapranov and others say that they have documented many examples of transcripts in which protein-coding exons from one part of the genome combine with exons from another

"We've come to the realization that the genome is full of overlapping transcripts."
— Phillip Kapranov



Spools of DNA (above) still harbour surprises, with one protein-coding gene often overlapping the next.

Hard to track

1860s After playing with pea plants, Austrian monk Gregor Mendel defines the basic rules of inheritance. Traits are determined by discrete units that are passed from one generation to the next.



1909 Danish botanist Wilhelm Johanssen coins the word 'gene' for the unit associated with an inherited trait, although the physical basis remains unknown.

1910 Thomas Morgan's work on fruitflies (right), shows that genes sit on chromosomes, leading to the idea of genes as beads on a string.



1941 George Beadle and Edward Tatum introduce the concept that one gene makes one enzyme.

1944 Genes are made of DNA, find Oswald Avery (below), Colin MacLeod and Maclyn McCarty.

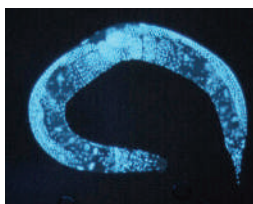


1953 James Watson and Francis Crick publish the chemical structure of DNA; the central dogma of molecular biology emerges in which information flows from DNA to RNA to protein.

1977 Richard Roberts and Phillip Sharp discover that genes can be split into segments, leading to the idea that one gene can make several proteins.

1993 The first microRNA is identified in the worm *Caenorhabditis elegans*.

2003 GeneSweep: Human geneticists come up with a definition for protein-coding genes in order to decide on a winner for a bet on the number of human genes. The winner is announced, but geneticists acknowledge that they don't know the true answer.



2006 The idea that human genes are one long continuum begins to emerge.

part that can be hundreds of thousands of bases away, with several other 'genes' in between. This continuum of genes might even spill over the boundaries of chromosomes: last year, Richard Flavell at Yale University School of Medicine in New Haven, Connecticut, documented human immune-system genes that seem to be controlled by regulatory regions from another chromosome⁶. "Discrete genes are starting to vanish," Guigo says. "We have a continuum of transcripts."

Slippery concept

The large transcriptional surveys suggest that a vast amount of the RNA manufactured by the mouse and human genomes do not code for proteins. Last year a consortium of researchers in Japan, for example, estimated that a whopping 63% of the mouse genome is transcribed^{7,8}; only 1–2% of the genome is thought to be spanned by sequences that contain everyday exons.

The discovery of RNA sequences that aren't just intermediates between the DNA and the protein-making machinery is not new in itself; the cell's protein-building apparatus requires a number of RNA molecules as well as proteins to operate. But the finding of 'microRNAs' and other RNA molecules now known to be vital in controlling many cellular processes in plants and animals, and the newly revealed ferment of RNA transcription, contributes to the view that RNA actively processes and carries out the instructions in the genome.

Perhaps the regions that make non-coding RNA should also carry the status of genes, if not the name itself. "I think it's time for people to take a deep breath and step back," says molecular biologist John Mattick of the University of Queensland in Brisbane, Australia. "A lot of the information in the system is being transacted by RNA."

Although functions have been identified for several RNA molecules, the crux of the debate now is the extent to which all the extra RNA plays a part. It is conceivable that it is easier to overtranscribe and ignore the rubbish than to invest in systems that produce only what is needed. A study from last year, however, hints that at least some of the mass of RNAs is doing something useful.

Working at the Genomics Institute of the Novartis Research Foundation in San Diego, California, John Hogenesch and his co-workers systematically quenched the activity of more than 500 non-coding RNAs in human cells and found that eight were involved in cell signalling and growth⁹.

But Hogenesch, and many other scientists, remain convinced that non-coding RNAs are much less important, functionally, than those that describe proteins; in the past,

when scientists have searched for the genetic basis of a disease or other characteristic they have overwhelmingly found the underlying mutation to be in a protein-coding gene rather than in another region. "The preponderance of evidence suggests that protein-coding genes will hold their own when the day is over," Hogenesch says.

Some of the recent discoveries — that the human genome makes a continuum of transcripts and that cells produce masses of non-coding RNA molecules — have not posed much of a problem to people outside the world of molecular biology. Population geneticists can examine how a trait is passed down and evolves regardless of the precise molecular mechanism that underlies it. For example, geneticists can build models showing how a mutation is inherited whether it affects a protein, a non-coding RNA or a regulatory region. "I don't actually care if it's making a protein or not," says Hurst. "The equations are still the same."

But the same can't be said for studies revealing so-called extragenomic modes of inheritance. In recent years, many investigators have focused on epigenetic inheritance, in which information is passed from parent to offspring independent of the DNA sequence. And this week in *Nature* (see page 469), Minoo Rassoulzadegan's team at the French National Institute for Health and Medical Research (INSERM) in Nice, France, reports that RNA may sometimes be complicating traditional models of inheritance.

In mice, mutations in the *Kit* gene cause white patches on the tail and feet; if a mouse has one normal *Kit* gene and one mutated one it will have the spots. The odd thing is

"A lot of the information is being transacted by RNA."

— John Mattick

that some of the offspring of such mice, who inherit two normal *Kit* genes, still have the white tail. The French group suggest that the mutant *Kit* gene manufactures abnormal

RNA molecules, which accumulate in sperm and pass into the egg. These bits of RNA somehow silence the normal *Kit* gene in the next generation and subsequent ones, producing the spotted-tail effect. "We are convinced that it's a more general phenomenon," says co-author François Cuzin.

If this is strange, the work reported last year¹ on the cress plant *Arabidopsis* by Robert Pruitt and his colleagues at Purdue University in West Lafayette, Indiana, is even stranger. Here the gene involved is called *HOTHEAD*. Pruitt and his co-workers' analysis shows that some plants do not carry the mutant version of *HOTHEAD* that their parents possessed. These plants had replaced the abnormal DNA sequence with the regular code possessed by earlier generations. "It's like, whoa, this changes everything," Pruitt says. "It definitely changes my view of inheritance."

Pruitt is now working to explain how the



Back-up copies: mutant DNA in the cress plant may be 'corrected' by inherited RNA.

plant could perform such a feat. One idea is that they carry a back-up copy of their grandparents' genetic information encoded in RNA that is passed into seeds along with the regular DNA and is then used as a template to 'correct' certain genes. Conceivably, Pruitt says, some of the mystery non-coding transcripts could be responsible. "I think there's something being inherited outside what we think of as the conventional DNA genome."

Changing views

The implications of such findings for our understanding of evolution have yet to be figured out. But research into the role of RNA as a carrier of information across generations

promises to enrich — and complicate — the notion of a gene yet further.

Leaving aside the can of worms that studies on epigenetics are beginning to open up, does it matter that many scientists not directly concerned with molecular mechanisms continue to think of genetics in simpler terms? Some geneticists say yes. They worry that researchers working with an oversimplistic idea of the gene could discard important results that don't fit. A medical researcher, for example, might gloss over the many different transcripts generated by a sequence at one location. And the lack of a clear idea of what a gene is might also hinder collaboration. "I find it sometimes very difficult to tell what some-

one means when they talk about genes because we don't share the same definition," says developmental geneticist William Gelbert of Harvard University in Cambridge, Massachusetts.

Without a clear definition of a gene, life is also difficult for bioinformaticians who want to use computer programs to spot landmark sequences in DNA that signal where one gene ends and the next begins. But reaching a consensus over the definition is virtually impossible, as Karen Eilbeck can attest. Eilbeck, who works at the University of California in Berkeley, is a coordinator of the Sequence Ontology consortium. This defines labels for landmarks within genetic-sequence databases of organisms, such as the mouse and fly, so that the databases can be more easily compared. The consortium tries, for example, to decide whether a protein-coding sequence should always include the triplet of DNA bases that mark its end.

Eilbeck says that it took 25 scientists the better part of two days to reach a definition of a gene that they could all work with. "We had several meetings that went on for hours and everyone screamed at each other," she says. The group finally settled on a loose definition that could accommodate everyone's demands. (Since you ask: "A locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions.")

Rather than striving to reach a single definition — and coming to blows in the process — most geneticists are instead incorporating less ambiguous words into their vocabulary such as transcripts and exons. When it is used, the word 'gene' is frequently preceded by 'protein-coding' or another descriptor. "We almost have to add an adjective every time we use that noun," says Francis Collins, director of the National Human Genome Research Institute at the National Institutes of Health in Bethesda, Maryland.

But however much geneticists struggle to pin down the elusive gene, it is precisely its ambiguous nature that fuels their continued curiosity. "It's ever more fascinating," says Whitehead's Young. Some things, it seems, are not best portrayed by a crude four-letter word. ■

Helen Pearson is a reporter working for *Nature* in New York.

Muddling over genes

Science philosophers Karola Stotz, at Indiana University in Bloomington, and Paul Griffiths, now at the University of Queensland in Australia, are attempting to measure the extent of working biologists' bewilderment over genes.

They collected together 14 weird and wonderful (but real) genetic arrangements and asked biologists to decide whether each represents one, or more than one, gene.

One is a DNA segment that uses some of the same protein-coding sequences to manufacture two entirely different proteins with distinct functions. In another, one 'gene' is nestled within the non-protein coding intron of another. Another protein is assembled when four different RNA molecules, made from DNA scattered over 40,000 base pairs, are assembled into one transcript.

Confused? So were the

500 biologists who completed the questionnaire. Stotz and Griffiths found that 60% are typically sure of one answer, and 40% are confident of another. Hardly any confess that they don't know.

Stotz wants to examine whether scientists working in separate disciplines tend to view the situations in different lights. "It will be interesting to know if there is some order to the confusion," Stotz says. **H.P.**

1. Lolle, S. J., Victor, J. L., Young, J. M. & Pruitt, R. E. *Nature* **434**, 505–509 (2005).
2. Rassoulzadegan, M. *et al. Nature* **441**, 469–474 (2006).
3. Cheng J. *et al. Science* **308**, 1149–1154 (2005).
4. Parra, G. *et al. Genome Res.* **16**, 37–44 (2006).
5. Akiva, P. *et al. Genome Res.* **16**, 30–36 (2006).
6. Spiliarakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. & Flavell, R. A. *Nature* **435**, 637–645 (2005).
7. FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) *Science* **309**, 1559–1563 (2005).
8. RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) and the FANTOM Consortium *Science* **309**, 1564–1566 (2005).
9. Willingham, A. T. *et al. Science* **309**, 1570–1573 (2005).

THE DARK SIDE OF THE SUN

The Sun occasionally hurls streams of particles towards Earth, where they can wreak havoc with satellites. Predicting these solar storms is hard, but some physicists believe we're about to face the biggest bout of solar flares in years. **Stuart Clark** reports.

Halloween is supposed to be a time of weird phenomena and spooky events. But by any standards what happened in late October 2003 was unusual. Telecommunications around the world were disrupted, half of NASA's satellites malfunctioned, 50,000 people in Sweden were left without electricity, and the global airline industry lost millions of dollars.

The link between these events was not supernatural, it was something far more familiar: the Sun. The chaos was caused as our star went through one of the more active moments in its 11-year cycle. And according to some predictions, what happened that October is nothing compared with what is going to occur in five or six years' time.

"Solar activity in the next cycle could be more of a problem than ever," warns Peter Gilman, a physicist at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado. And it's not just satellites and telecommunications that face problems. Some researchers claim that the Sun's behaviour affects Earth's atmosphere — in particular influencing cloud formation. This claim has attracted global-warming sceptics, who argue that the Sun has greater influence than human activities on our changing climate.

The Sun's 11-year cycle is driven by its magnetic field, and generates a flow of charged particles known as the solar wind. At the quieter parts of the cycle, activity is fairly low and the solar wind is reasonably uniform. But at the 'solar maximum', sunspots — dark patches caused by the magnetic field twisting at the surface — appear on the Sun's face. Huge solar flares explode above these spots causing turbulence in the solar wind and sending streams of charged particles hurtling through space.

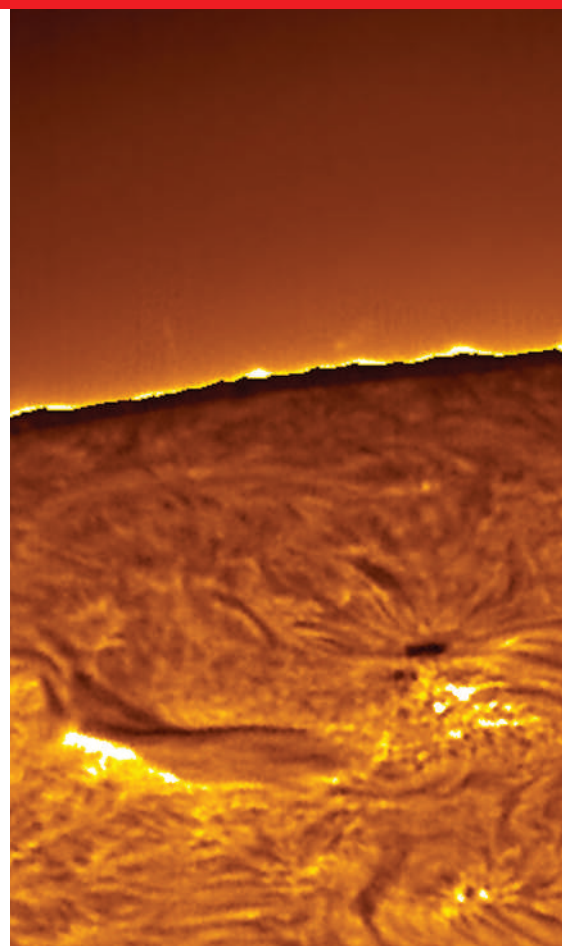
The most recent solar cycle was fairly moderate when measured by the number of sunspots (see graphic). Yet in late

October 2003, three years after the cycle's peak, two monstrous sunspots appeared, each more than ten times the diameter of Earth. Both were in a state of almost constant eruption, spewing out billions of tonnes of electrically charged particles.

These were the particles that caused such havoc when they hit Earth's atmosphere. The global maritime emergency call system blacked-out, contact was lost with expeditions on Mount Everest, and the accuracy of the global positioning system was impaired. As well as NASA's satellite malfunctions, the Japanese lost contact with one of their weather satellites altogether. The cost to the airline industry arose as planes were re-routed to lower altitudes, congesting the airways and burning more fuel.

Lucky escape

The sunspots bombarded Earth, on and off, for two weeks as the Sun's rotation carried them across its face. On 4 November, as the second sunspot was about to be lost from sight, it let loose another tremendous explosion. Solar physicists calculated that it was one of the largest solar flares in recorded history¹. By sheer luck it exploded into deep space, catching Earth only in the side wash. Those who saw it breathed a sigh of relief and won-



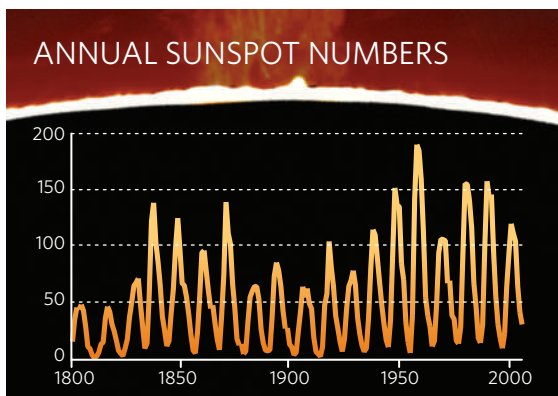
dered what the damage might have been if such a flare had exploded facing Earth. If the latest prediction comes true for the next solar cycle, we may yet find out.

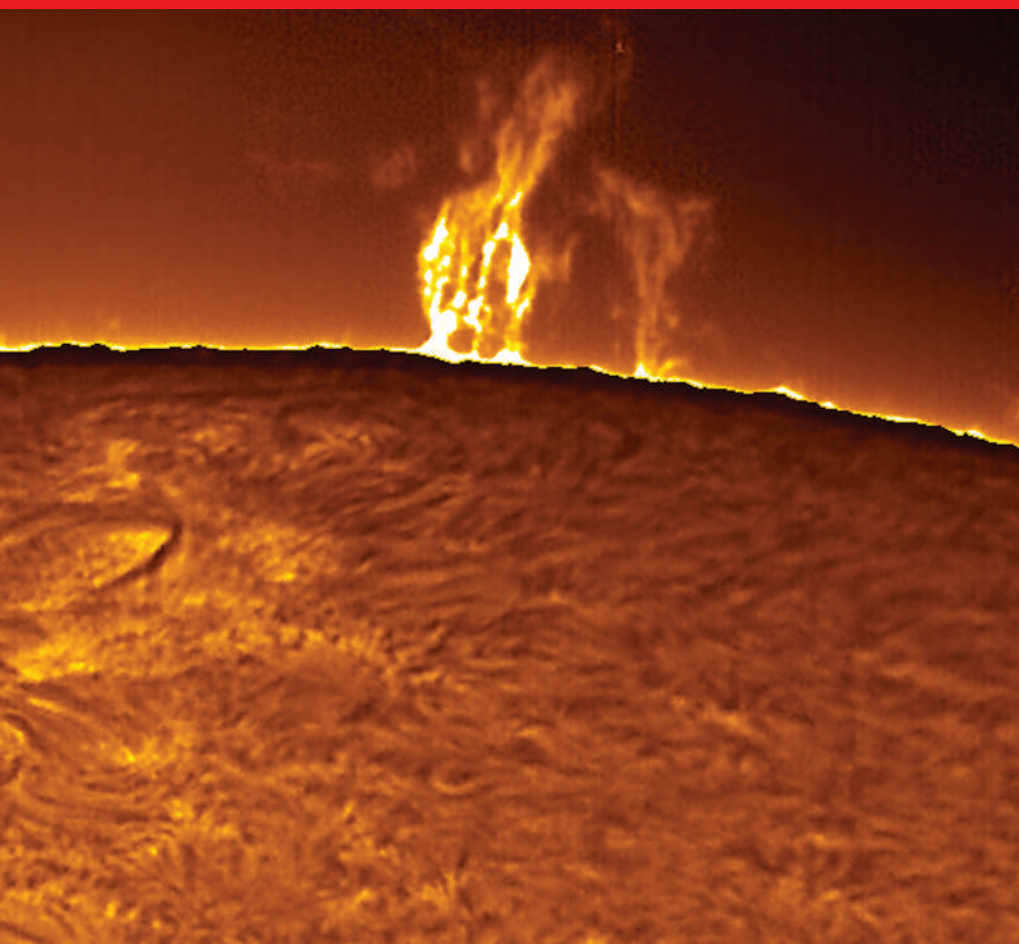
Predicting the timing and strength of such solar eruptions is clearly important, but it is hampered by the fact that scientists know relatively little about the Sun's inner workings. So to coincide with the start of the next solar cycle, the largest coordinated study of the Sun will be launched next year. Known as the International Heliophysical Year (IHY), the initiative hopes to build awareness of the Sun's possible influence on Earth's climate and to bring researchers from different disciplines together to study solar activity.

Currently, the Sun is at a solar minimum, and most predictions suggest that the next solar maximum in five or six years' time will be weak. But the most recent forecast, the first to be based on a completely physical model of the Sun, suggests otherwise.

This forecast has been generated by Mausumi Dikpati and her team at the NCAR². They have developed a computer simulation that mixes the Sun's internal magnetic dynamo with theories about how solar plasma circulates near the surface. And they have reached a sobering conclusion. "We expect between 30% and 50% more sunspots and solar activity than the cycle just ending," says Gilman, who is a member of Dikpati's team.

The last time solar activity occurred





Flaring up: the Sun's magnetic field causes a peak in sunspots and solar flares every 11 years.

on this scale was in 1958, when there was little technology in orbit. Now things are very different: Earth is surrounded by thousands of active satellites.

Satellite operators rely on predictions of solar activity to estimate the lifetime of space missions. The solar wind heats Earth's thin upper atmosphere, increasing atmospheric density and causing more drag. Gilman estimates that a 30% increase in activity will almost double the atmospheric density at an altitude of 300 km, affecting low-altitude satellites.

Mission planners looking ahead to 2012 may want to boost their spacecraft to higher orbits, or accept a shorter operational lifetime. Even above 800 km, where satellites are safe from atmospheric drag, other dangers remain. The solar wind can cause a build up in electrical charge, which then short-circuits and burns out sensitive equipment. This is the suspected fate of the Japanese Midori 2 satellite, lost during the 2003 flares. And as more satellites die in orbit, operators have to worry about dodging 'space junk'. In the aftermath of a large solar storm, the change in atmospheric drag can shift the orbit of space debris, endangering active satellites.

The Sun's influence over space hardware is only one aspect of the latest drive to understand the star. The possible effects of the solar cycle on our climate, especially cloud formation, are also receiving a lot of attention. A link

between the two was suggested in 1997, when meteorologists Henrik Svensmark and Eigil Friis-Christensen, both at the Danish Meteorology Institute in Copenhagen, analysed weather satellite records for 1979 to 1992. They found that during solar minima, Earth was 3% cloudier than at solar maxima³. They also noticed that the influx of high-energy particles reaching Earth from deep space,



**"We must now
let Mother
Nature tell us
who is right."
— Leif Svalgaard**

phenomena known as cosmic rays, was up to 25% higher at solar minima, hinting that they might seed cloud formation. The pair called their finding a "missing link in solar-climate relationships".

Climate sceptics who argue that human activities are not responsible for global warming have seized on these results. They claim it shows that the Sun is largely responsible for variations in our climate. So convinced are they that last year two Russian sceptics placed a \$10,000 bet that global temperatures will show an average fall for 2012–17 — on the assumption that the next solar cycle will be weak⁴.

But most proponents of the solar-climate link are proceeding more carefully. "We're not suggesting that all clouds are formed by solar activity, merely that the process might be modulated by solar activity," says Robert Bingham, a physicist at the Rutherford Appleton Laboratory in Didcot, UK. He is part of an international experiment known as CLOUD, or Cosmics Leaving Outdoor Droplets. This will use CERN's particle accelerator on the French-Swiss border to fire charged particles through a chamber holding gases to simulate Earth's atmosphere and determine whether 'clouds' are created.

Global network

To take advantage of the next solar cycle more directly, the United Nations is heading an initiative to install radio receivers in all 191 of its member states. For the first time, the upper atmosphere's response to the continual collision of solar radiation would be monitored on a global basis. Although space officially starts at an altitude of about 100 km, scientists know little about this region because it is difficult to study.

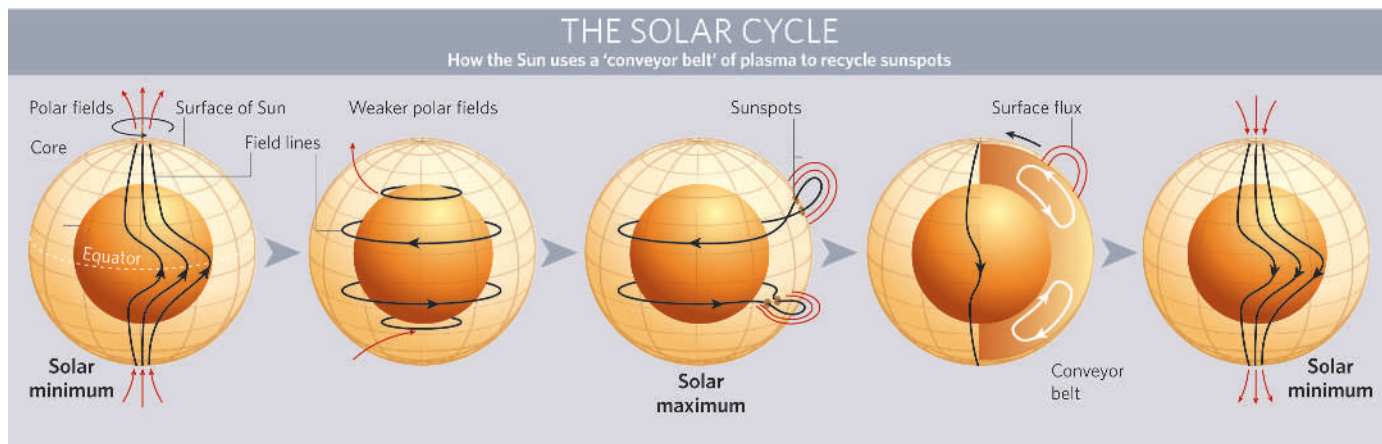
The UN project is one of the planned elements in the IHY. Although it has no dedicated research budget, the IHY has initiated a call-for-proposals aimed at making it easier for scientists from any discipline to gain access to solar instruments and data. "We are inviting ideas from the community," says Rutherford's Richard Harrison, the joint UK coordinator for the IHY.

Certainly 2007 will put at scientists' disposal the largest-ever fleet of space missions for studying solar-terrestrial interactions. A dozen spacecraft that track solar activity are already in orbit, and another three should launch this year, including the most sophisticated solar watchdog yet. NASA's Solar Terrestrial Relations Observatory (STEREO) consists of two nearly identical craft that will watch the Sun from different locations, one preceding Earth in its orbit and the other following behind. This will allow them to take stereoscopic images of the Sun and to track the three-dimensional structure of particle eruptions. In this way, STEREO might be able to supply advance notice of the speed and direction of eruptions as they head towards Earth.

Such information should help satellite operators respond to imminent dangers, but for proper planning they will need long-term forecasts of solar activity. Some researchers, such as Harrison, believe that scientists don't yet understand the Sun enough to make meaningful long-term predictions. Certainly, past forecasts have relied on tracking signposts of future solar activity, without worrying too

WWW.SUNGAZER.NET

M. SVALGAARD



much about the mechanisms behind them.

For example, in the 1970s, astronomers recognized that the build up of magnetism at the Sun's poles after the cycle has peaked has a bearing on the strength of the next cycle. Just last year, one of the pioneers of this method, Leif Svalgaard, used the Sun's polar magnetic field to predict that the next solar cycle will be the weakest for a century⁵. Other 'signpost' methods, such as those looking at the amount of 10.7-centimetre radio waves coming from the Sun or the number of bright patches near the Sun's poles, also forecast a weak cycle.

The only signpost method to predict a strong cycle comes from solar physicists David Hathaway and Robert Wilson at NASA's Marshall Space Flight Center in Huntsville, Alabama. In 2004, they noticed that a solar cycle's strength correlates to the number of sunspots two cycles before. Applying that rule of thumb to the next cycle, they have predicted strong activity in 2012 (ref. 6). Dikpati's model agrees with this forecast and, crucially, puts the reason for it on a physical footing.

In the past decade, physicists have discovered a vast conveyor belt of plasma on the Sun that seems to flow from the equator to the poles in each hemisphere at around 30–65 kilometres per hour. Sunspots are typically active for just a few weeks before fading from view, but their magnetic fields linger on. These weak fields are carried by the flow and accumulate at the poles before being submerged below the surface, where they presumably flow back towards the equator⁷.

Dikpati's work combines sunspot observations dating back to the 1900s with a computer simulation of the Sun's magnetic dynamo and the conveyor belt (see graphic). In the simulation, the conveyor belt sweeps along old sunspots,

submerging them at the poles. During the deep return flow, the Sun's rotation rejuvenates the old magnetic fields, creating new sunspots and fresh areas of solar activity.

It is the only prediction in which every step uses a physics-based computer model, which is why it is being taken seriously by solar physicists. "The solid physics of Dikpati's model is a high hurdle for the other techniques to get over," says Hathaway.

Solar memory

The key to Dikpati's forecast is how fast the Sun's conveyor belt runs. The deep return flow is unmeasurable but the model suggests that it is slower than the surface flow, perhaps just 5 kilometres per hour. If so, the return leg of the journey would take a couple of decades. "This shows that the Sun retains a memory of its magnetic field for

about 20 years," says Dikpati. So in her model, the Sun's activity is not based solely on the previous cycle's magnetic field but on the interplay with earlier cycles. In contrast, most 'signpost' prediction methods assume that the previous solar cycle immediately kicks off the activity of the next. "It is good for science

that the predictions are now diverging," says Svalgaard, although he disagrees with Dikpati's conclusions.

Solar physicists are now waiting to see if this physics-based forecast is right. And

they may not have to wait for the peak of activity in six years' time to find out. All methods predict only the average number of sunspots, but records show that large cycles have always begun early and raced to their peak. That means that the telltale signs of a large solar cycle should be evident within just three or four years from now.

"We must now let Mother Nature tell us who is right," says Svalgaard. But Dikpati and her team are refining their model to see whether it can predict features such as an early start. Either way, there will be plenty of sun watchers — from mission planners to climate sceptics — tracking the way the solar wind blows.

Stuart Clark is a freelance writer based in Hertfordshire, UK.



Sun spotters: NASA's twin STEREO satellites will watch the solar weather and give warning of approaching storms.

1. Tsurutani, B. T. *et al. Geophys. Res. Lett.* **32**, L03S09 (2005).
2. Dikpati, M., de Toma, G. & Gilman, P. A. *Geophys. Res. Lett.* **33**, L05102 (2006).
3. Svensmark, H. & Friis-Christensen, E. *J. Atmos. Solar-Terr. Phys.* **59**, 1225–1232 (1997).
4. Giles, J. *Nature* **436**, 897 (2005).
5. Svalgaard, L., Cliver, E. W. & Kamide, Y. *Geophys. Res. Lett.* **32**, L01104 (2005).
6. Hathaway, D. H. & Wilson, R. M. *Solar Phys.* **224**, 5–19 (2004).
7. Hathaway, D. H. *et al. Astrophys. J.* **589**, 665–670 (2003).

BUSINESS

Carbon market survives gas leaks

Searing volatility has led some to dismiss Europe's nascent emissions market as a farce — but it is still hanging in there. **Quirin Schiermeier** reports on the project's teething troubles.

On the closing Friday of Carbon Expo in Cologne earlier this month, a crowd of nervous emissions traders was seen swarming around the stand of a German consultancy firm.

The focus of their attention was an analyst who had managed to log on to a supposedly secure European Commission website. The site listed verified data on carbon dioxide emissions from thousands of large companies. The commission was due to release these market-sensitive data the following Monday, 15 May.

As reports spread that power producers and other energy-intensive European industries were sitting on unused allowances for 60 million tonnes of carbon dioxide emissions, the price of the permits that allow for extra emissions dropped by nearly a third. Traders' reactions to the leaking of the data varied from "unbelievable" to "a complete farce". Only the arrival of the weekend, they said, saved the market from a complete meltdown.

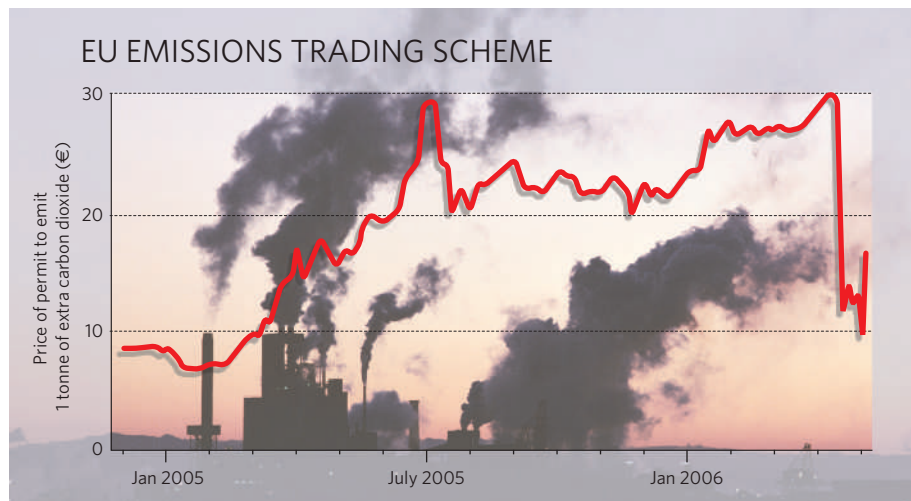
When the commission officially announced on 15 May that large industrial firms in the European Union (EU) were in fact just 44 million tonnes under the permitted limit for 2005, the market decided that the crisis wasn't quite so bad. It settled somewhat, at about half its April peak.

The commission later apologized for the "technical error". But the incident summed up the immaturity of the EU Emissions Trading Scheme (ETS), a cap-and-trade system introduced in January 2005 with the goal of cutting greenhouse gases. The world looks to the ETS as a possible model for a global trading system, and is watching its progress with interest.

Accidental emissions

The leak wasn't the first of its kind: late last month, prices of emission allowances crashed from €31 (US\$40) to around €12, following leaks of emissions data for France, the Netherlands, Sweden, Spain, Belgium and the Czech Republic. Most of these countries reported significantly lower emissions than had been foreseen by their respective national plans.

Not all countries have come in below target: British companies emitted 33 million tonnes more carbon dioxide than they were allowed. The excess will put particular pressure on emitters in export-dependent sectors such as the steel industry, which cannot readily hand on the



cost of buying permits to their customers. The largest emitters — electricity generators — probably will hand on such costs, but they're not happy about it: five British firms are suing the European Commission, claiming that the UK allocation is too small.

Across the EU, however, over-allocation of emissions rights in the first trading period seems to be the biggest threat to the ETS.

"If the price of carbon is too low, it will hinder investment in cleaner technologies," says Martina Priebe, who watches the ETS for the International Emissions Trading Association in Geneva, Switzerland. "In the worst case the whole cap-and-trade exercise will have no effect at all on investment decisions."

For the system to work properly, the total number of allowances allocated to industry needs to be smaller than the actual emissions, agrees Barbara Helfferich, a spokeswoman for the European Commission. But she says it was difficult to estimate the right baseline because data on past emissions levels didn't exist. Instead, the commission used data on past energy consumption and projected economic growth rates to estimate allowances for the 11,500 installations covered by the ETS.

The 25 EU member states must submit their national allocations for the second trading period, 2008–12, by the end of June. These can now take into account verified data and

will be adjusted accordingly, says Helfferich.

"The commission acted on the cautious side," says Guy Turner, director of London-based New Carbon Finance. "This is understandable because they knew that excessively high prices would have been disastrous."

Volatility may be a problem for investors but it is typical for young environmental commodity markets, such as the one that successfully controls sulphur dioxide emissions in the United States. And although the ETS market may have fallen, it is still functioning.

"Emissions trading works," says Peter Kreuzberg, head of trading at the German energy company RWE. He notes that high market volume and liquidity have been achieved quickly. But he adds that the ETS reporting system needs to be strengthened and its allocation process made more transparent.

These flaws must be addressed if the scheme is to help the EU meet its target of reducing industrial carbon dioxide emissions to 6% below current levels before 2010. "The performance of the ETS stacks up well against most criteria," says Turner. "The only thing it has not yet done is reduce emissions."

"The system is not a failure despite what has happened in the past few weeks," says Priebe. The commission can't control the market's value — but, she says, it needs to do far more to control the quality of data that underpin it.

"The only thing the trading scheme has not yet done is reduce emissions."

Climate: open review may ease acceptance of report

SIR — As executive director of the Office of the US Global Change Research Program from 1993 to 1997, I was responsible in 1995 for urging adoption of the national review process of the IPCC report that is questioned in your News story “US posts sensitive climate report for public comment” (*Nature* 441, 6–7; 2006).

The IPCC has always used a multi-stage review process. The first review is by nominated technical experts. In the second review, however, copies of the report are also sent to governments and all non-governmental organizations registered with the IPCC. In this round, the governments can each choose their own review process. What led to the US decision for open review was a query from one industry association that was interested in the environment but had not been sending a representative to IPCC plenary meetings. It asked why it was denied a chance to review the report when similar groups with a clear advocacy position were not. We could think of no acceptable answer. If all the world's nations and hundreds of scientists and non-governmental groups could review the report, how could we say some groups could not?

In response, at my urging, the US government agreed to a process whereby notice of the report's availability would be published in the *Federal Register*, calling for comments. This has been the practice ever since, and initial fears that it would be abused, for example, by large numbers of coordinated pleadings, have not materialized.

We did insist that those submitting comments identify themselves: a step we hoped would encourage careful and responsible presentation of comments. But we decided that what would matter was the merit of the comment, not its origin. Each comment was evaluated by an interagency team of government programme managers with interests and expertise in the area under consideration. In general, most comments were accepted: industry reviewers often had special insight into issues relating to emissions and approaches to mitigation; environmental-group reviewers had views on ecosystem behaviour and related pollution issues; and public reviewers had views on the clarity of presentation, or the significance or uncertainty of the findings. There were some off-the-mark and gratuitous comments, but not many, and these were filtered out by the interagency review process.

My impression was that the process worked well through the second and third IPCC assessments, and that the comments submitted helped to improve the overall quality of the assessments. There were instances of a strong difference of opinion within the expert community. But in all cases,

through communication between authors and reviewers, we were able to facilitate a positive outcome. There was a sense that all legitimate perspectives (those written up in peer-reviewed or other equivalent form) had been fairly considered.

I have not been part of the government's planning or conduct of this review round. I am impressed, however, by their commitment to continuing the open approach. Let's not fear getting comments; in my opinion, promoting wide involvement in the review process, where questions can be clarified, has much greater potential for winning general acceptance of the findings than springing results on uninformed parties through some media blitz at the end of the process.

Michael MacCracken

Climate Institute, 1785 Massachusetts Avenue NW, Washington DC 20036, USA

Climate: US has always made IPCC drafts available

SIR — Your News story “US posts sensitive climate report for public comment” (*Nature* 441, 6–7; 2006) implies that the United States has departed from its traditional approach in reviewing draft reports from the Intergovernmental Panel on Climate Change (IPCC). In fact, US procedures, first published in the *Federal Register* in 1995, reflect our longstanding commitment to open IPCC reviews.

Under the 1995 procedures, we provided paper copies of IPCC draft reports to any individual upon request. In 2000, individuals were also asked to provide information on their “qualifications and general area of expertise ... to review specific parts of the report”, although to the best of my knowledge drafts were made available to all who submitted requests. Since September 2004, after legal counsel advised that requiring individuals to provide such information was inconsistent with federal information-dissemination guidelines, we have made IPCC drafts available on the US Climate Change Science Program website (www.climate-science.gov/Library/ipcc); this prominently displays IPCC instructions that drafts not be cited, quoted or distributed. As far as we were aware, everyone accessing this website had abided by the IPCC instructions, until *Nature* published elements of the Fourth Assessment Report.

Our approach reflects the view that it is neither possible nor appropriate for the government to decide which individuals are expert enough to review the report. Comments on the draft will be reviewed on their merits by prominent scientists during both US and international IPCC reviews. To ensure objectivity, submitted comments will be supplied to federal programme managers

and scientists without attribution and then evaluated on scientific and technical grounds before being sent to the IPCC.

The United States is committed to an effective review of the IPCC Fourth Assessment Report, and we intend to provide comments reflecting the extensive expertise of the US scientific community.

Harlan L. Watson

US Department of State, 2201 C Street NW, Washington DC 20520, USA

HIV denialists ignore large gap in the study they cite

SIR — Valendar Turner, in Correspondence (“HIV drug remains unproven without placebo trial” *Nature* 434, 137; 2005), argues that there is no evidence for antiretrovirals reducing the transmission of HIV from mother to child. He points out that HIV transmission in people taking the antiretroviral drug nevirapine was 13.1% in the HIVNET 012 study in Uganda, whereas only 12% of women in a Rwandan study were found to have transmitted HIV to their babies in the absence of antiretroviral treatment.

Despite a rebuttal in Correspondence by the authors of the Ugandan study, Brooks Jackson and Thomas Fleming (“A drug is effective if better than a harmless control” *Nature* 434, 1067; 2005), Turner's letter continues to be cited by AIDS denialists (for example, C. Farber *Harper's Magazine* 37–52; March 2006).

The Rwandan study referred to by Turner enrolled 561 pregnant women, of whom 286 were HIV-positive. Of the children born to HIV-positive mothers, 158 were tested for HIV and 19 (12%, as Turner states) were found to be HIV-positive. Why were only 158 children assessed? The answer, conveniently ignored by the denialists, is that follow-up was interrupted by the events of the Rwandan civil war (J. Ladner *et al.* *J. Acquir. Immun. Def. Syndr. Hum. Retrovirol.* 18, 293–298; 1998). Given that this interruption was sufficiently lengthy for many HIV-positive children and their mothers to die of AIDS in the interim, the surviving sample of the initial cohort cannot be regarded as representative. The actual figure for HIV transmission was almost certainly much higher. Failing to acknowledge this important caveat to the study appears to us to be inconsistent with accepted academic standards.

Nathan Geffen*, Nicoli Nattrass†, Glenda Gray‡

*Treatment Action Campaign, 34 Main Road, Muizenberg 7945, South Africa

†AIDS and Society Research Unit, University of Cape Town, Private Bag, Rondebosch 7701, Cape Town, South Africa

‡Perinatal HIV Research Unit, University of Witwatersrand, PO Box 114, Diepkloof 1864, South Africa

BOOKS & ARTS

Building on failure

Working out why something doesn't work is a good starting point for improving the design.

Success Through Failure: The Paradox of Design

by Henry Petroski

Princeton University Press: 2006. 240 pp.
\$22.95, £14.95

J. M. Ottino

All the products of human activity we see around us, including the layout of the magazine or website you are reading now, were invented and designed. Design is so prevalent that to some degree most of us are designers. This does not mean we are good at it, or that we have ever given much thought to what it means to produce a successful design.

Not that success goes together with design in Henry Petroski's world. The title of his latest book makes this clear: success is linked to failure. Like theory and truth in Karl Popper's world, a design can never be proved to be a success; success is provisional, it is just the absence of failure. One can hardly improve on the author's words: "Things that succeed teach us little beyond the fact that they have been successful; things that fail provide incontrovertible evidence that the limits of design have been exceeded. Emulating success risks failure; studying failure increases our chances of success. The simple principle that is seldom explicitly stated is that the most successful designs are based on the best and most complete assumptions about failure."

Even though it is dangerous to put the words so close together, there is a clear flavour of evolution — in terms of success riding on the back of failure — to the theme of this design-centred book.

This is but a slice of the design picture, and it is intrinsically unfair to judge a book by what it does not cover. But in order to place the book on a broader canvas, let me mention two things.

The first is that building on the back of failure may sound easy but it is not. Failure is often hard to pinpoint, which is why forensic engineering has emerged as a distinctive discipline. In our rapidly evolving technological world, there can be an almost infinite array of possible design states, and not all can be tested. Plane crashes provide some of the most visible examples, in terms of life lost, cost and the politically charged assignment of cause. The cause of the explosion of TWA flight 800 over Long Island on 17 July 1996 was initially



Safe as houses: the best way to ensure the safety of skyscrapers is to learn from previous failures.

believed to be a terrorist bomb or a missile. But a meticulous reconstruction of recovered pieces, aided by sophisticated laboratory and mathematical analyses, indicated that the likely cause of the explosion was a design defect of the centre fuel tank.

The second is the sheer creativity in design, which makes one wonder if evolution-driven design can be taken too far. True, everything comes from somewhere. Retrospective analyses show that even artistic masterpieces do not suddenly appear from thin air — the final painting is deceptive when presented without context. But some designs are much more than tweaks to previous art, and seemingly appear from nowhere; think of the iPod, the computer mouse, the BlackBerry and even the Internet.

But there is no question that failures drive design. "Fail early, fail often" is a mantra in product development. The examples in this book take us through the evolution of bridges, the world's tallest buildings, PowerPoint, the space-shuttle disasters, the collapse of the World Trade Center, flag design and constitutional engineering. Most of the examples are from the worlds of civil and mechanical engineering. The chapters on bridge evolution provide some beautiful and clear illustrations, and there is an intriguing case of a thirty-

year cycle of failure yielding a prediction.

The book poses questions that should have occurred to us all but probably didn't. How does one test ever more complicated systems against all possible conditions? How do we test civil engineering projects, such as dams, tunnels, buildings and bridges, "whose scale is so large, whose cost is so great, and whose design is so specific to the site that the structure is unique"?

'Design think' is now at the centre of discussions about technological leadership. It connects with innovation, a topic arising in nearly a dozen high-profile reports in the past two years. It is also emerging as a central theme providing a competitive edge in the strategy of corporations. There is no recipe to good design think; it is non-algorithmic and, as with many creative activities, it pays to practise.

Recent books have brought economics to the masses, and there now seems to be a trend to do the same with design. This is a good thing and this book, like several earlier ones by Petroski, is part of this very welcome trend. *Success Through Failure* is insightful and accessible. I hope it is widely read. ■

J. M. Ottino is at the R. R. McCormick School of Engineering and Applied Sciences, Northwestern University, Evanston, Illinois 60208-3100, USA.

A mammoth murder mystery

Twilight of the Mammoths: Ice Age Extinctions and the Rewilding of America
by Paul S. Martin

University of California Press: 2005. 269 pp.
\$29.95, £18.95

Alan B. Shabel

In 1877, Richard Owen suggested that the great Pleistocene fossil mammals of Australia had been driven to extinction by “the hostile agency of man”. Seven years later, C. S. Wilkinson replied that a reduction in rainfall, leading to the impoverishment of a once rich flora, was a more likely trigger for the loss of these beasts. And so it has always been in the study of late Quaternary extinctions: arguments for anthropogenic causes are countered by arguments in favour of climate change.

It is widely accepted that in the past 50,000 years, the world has lost a diverse array of reptiles, birds and mammals. A striking feature of this mass extinction event is that it disproportionately affected large-bodied vertebrates (more than 45 kg). With few exceptions, small vertebrates, plants and insects did not suffer dramatic losses. Another striking feature is the extinction event's geographic asynchrony: the extinctions occurred first in Australia (about 46,000 years ago) and much later in the Americas (13,000 years ago), the West Indies (5,000 years ago), Madagascar (2,300 years ago) and New Zealand (500 years ago). On each land mass, the extinctions occurred soon after the immigration of *Homo sapiens*, and the timing of the extinctions does not correlate with anomalous changes in climate.

In the 1960s, Paul Martin sparked the modern study of the late Quaternary extinctions with his ‘overkill’ hypothesis, which argues forcefully for the role of human hunting in the extinction event. In his new book, *Twilight of the Mammoths*, Martin synthesizes fifty years of his research on the subject in a palaeontological memoir. He takes an extreme position, in two respects. First, he argues that humans



M. LONG/NHMPL

Was climate change or hunting by humans to blame for the demise of large mammals in the Pleistocene?

wiped out the large vertebrates through direct predation pressure; although he acknowledges the potential role of less direct human influences, such as pathogen propagation and artificial fires, he focuses primarily on hunting. Second, he posits that the extinctions occurred extremely quickly; for example, he thinks the whole of North America could have lost its megafauna in “a few tens or hundreds of years”.

It is convenient that Martin takes such an extreme position because it renders his model testable. If it can be shown that humans co-existed with megafauna for a thousand years or more, his blitzkrieg version of the overkill model would be in trouble. In most cases, such as in Australia and South America, the palaeontological record is not known in sufficient detail to determine the precise dates of human arrival and megafaunal extinction. But the situation is better in North America and on the large oceanic islands, where the timing of the two phenomena correlate very closely. However, if humans were responsible for such large-scale devastation, where are the archaeological kill sites? In North America, only a dozen or so late Quaternary sites have been found with clear evidence of butchery (and almost all of these are mammoth), and there are no known sites with artistic depictions of humans interacting with any members of the extinct megafauna.

To most observers, the paucity of kill sites constitutes strong evidence against the overkill model. Martin, however, argues that this absence of evidence actually supports his position: if the extinction event was extremely rapid, the devastation would fail to register in the archaeological record. Overkill seems to have become an *idée fixe* for Martin, and his logic becomes circular at times. For example, he writes: “I have grave doubts about the existence

of a widespread and biologically effective human population in the Americas before 13,000 years ago precisely because large, slow-moving, eminently huntable animals such as ground sloths continued to occupy their favorite dung caves in North and South America as late as they did.” In other words, he takes the presence of megafauna to signal the absence of humans, because the two could not possibly have coexisted. Martin is certainly more effective when he is working to test his model than when he is assuming it.

As if his overkill hypothesis were not controversial enough, Martin finishes the book with a call to restore lost megafauna through the introduction of closely related species, or ecological proxies, into the ‘blighted’ areas. In North America, he declares, we ought to restore not only bison and brown bears, but also proboscideans, camels and lions. If giant ground sloths are not available, we could use rhinos as surrogates. Martin recognizes that his solution is not perfect (and that in some cases the political hurdles may be insurmountable), but he argues that this attempt at restoration is “infinitely better than continual attrition”.

Whatever scientists may think of Martin's call for ‘resurrection ecology’, it must be agreed that his perspective fundamentally challenges our conception of what is ‘natural’. Many ecologists seek to study systems without a contemporary human presence so they can better assess the pure dynamics of nature. Martin's work impresses on us the fact that Earth's terrestrial ecosystems were profoundly restructured long before European colonialism and industrialization. That, at least, no longer seems controversial. ■

Alan B. Shabel is in the Department of Integrative Biology, University of California, Berkeley, Berkeley, California 94720-3140, USA.

ELECTRIFYING BOOK WINS AVENTIS PRIZE

David Bodanis has won this year's Aventis Prize for science books. His book, *Electric Universe: How Electricity Switched on the Modern World* published by Little, Brown, was described by the judges as “a very good read” that “opens up a universe of facts that would scarcely be credible in an imaginary tale”.

The junior prize went to Kate Petty, Jennie Maizels and Corina Fletcher for their book *The Global Garden*, inspired by the Eden Project and published by Eden Books (Transworld).

The winners received their awards at a ceremony on 16 May 2006 at the Royal Society in London.

Everyone hates a know-all

The Last Man Who Knew Everything: Thomas Young, the Anonymous Polymath Who Proved Newton Wrong, Explained How We See, Cured the Sick, and Deciphered the Rosetta Stone, Among Other Feats of Genius

by Andrew Robinson

Pi Press: 2006. 288 pp. £17.99

David Philip Miller

Know-alls can be maddening. Polymathy, however fascinating, may be a vice as much as a virtue in this age of specialization. In his book *The Last Man Who Knew Everything*, Andrew Robinson attempts only an introduction to the life of the polymath Thomas Young (1773–1829) — a full biography of ‘Phenomenon Young’, as he became known, would have required a team of specialists.

Robinson aims to introduce Young’s life and achievements to a broad audience, and his readability quotient is high. The book is a superior example of the recent flood of popular works on the history of science and technology no doubt inspired by the success of Dava Sobel’s *Longitude* (Walker, 1995). Robinson frequently tells us how remarkable Young’s achievements were and defends him from Young’s contemporaries who found him obscure, difficult and, yes, maddening. Robinson’s parting shot is to compare Young to Leonardo Da Vinci. However, the hyperbole is balanced by Robinson’s genuine interest in why Young was not universally admired, and why he has received a mixed press from specialist historians. The meditations on specialization this induces are interesting but, in my view, fall short of explaining the negative reactions to Young.

Young came from a Quaker family who recognized his precocity early. Tutors and schoolmasters fed his hunger for languages — by his teens he spoke Greek and Latin, and had a working knowledge of Hebrew, Chaldee, Syriac, Samaritan and Persian. He taught himself mathematics and natural philosophy before attending medical lectures in London and studying at the universities of Edinburgh, Göttingen and Cambridge. He became a fellow of the Royal Society at the age of just 21.

Best known for his work on the interference of light and the development of a wave theory in opposition to the prevailing newtonian corpuscular ideas, Young also made crucial contributions to understanding the physiology of the eye and vision. In 1814 he began to decipher the demotic and hieroglyphic inscriptions of the Rosetta Stone, but his important contributions were overshadowed by Jean-François Champollion’s subsequent efforts. Beyond this, Young’s versatility knew few bounds, as shown by his numerous contributions to the *Encyclopaedia Britannica* and his

published lectures on natural philosophy and the mechanical arts. Young’s profession was medicine but his practice was not very successful; an inheritance and other sources of income meant that this was not a problem, however. Having abandoned Quakerism, he moved in high social circles in London. He died at the age of 56 from pulmonary troubles and, Robinson speculates, intellectual exhaustion.

So why the lack of universal admiration? Personality was an issue: Young comes across as self-contained but self-confident and unlikely to suffer fools. Robinson is right, also, in arguing that Young was a victim of a specializing age, of a mixture of envy and disbelief that he could successfully traverse so many fields. However, Robinson fails to consider sufficiently the social and intellectual context of the learned world of the metropolis in which Young grew to maturity. In the 1820s, new cultural patterns were pitched against the old. In the old pattern, Young’s range of activities was not unusual, except in the success that he enjoyed. He was one of a type — a medical doctor, versed in classics and natural philosophy, and interested in antiquities. In fact these were the very people that specializing (but hardly

specialist) young Turks fought against in the 1820s as they challenged the existing power structures of science in the Royal Society and elsewhere. This can explain much of the hostility towards Young when he was foreign secretary of the Royal Society, secretary of the Board of Longitude, and editor of *The Nautical Almanac*. Pushy young Cambridge scholars welcomed Young’s advocacy of wave theory but were exasperated by his mathematical obscurity. Along with their astronomical allies in London, they sought to reform (not abolish, as Robinson has it) the key astronomical institutions in which Young represented, culturally and temperamentally, the old guard.

Greater awareness of this institutional and cultural context would have lifted the narrative above being an encomium to Young, insightfully placed as it is against the backdrop of the generic problem of specialization. Discerning exactly what specialization meant in early nineteenth-century Britain is important to understanding why this particular know-all was so maddening as well as an object of justified admiration. Young was truly remarkable, but was also a creature of his times, not a Colossus astride them.

David Philip Miller is associate professor of history and philosophy of science at the University of New South Wales, Sydney, New South Wales 2052, Australia.

Life, the Universe and entropy

Decoding the Universe: How the New Science of Information is Explaining Everything in the Cosmos, From our Brains to Black Holes

by Charles Seife

Viking: 2006. 288 pp. \$24.95

Robert J. McEliece

Near the end of his new book *Decoding the Universe*, Charles Seife argues with a straight face that there exists a universe parallel to ours that is populated by a race of superintelligent octopuses. The funny thing is, I believed him.

If Douglas Adams had not already used it, a more appropriate title would have been *Life, the Universe and Everything*, for that is in fact just what this book is about. Seife tells us that the modern science of ‘information theory’ (I’ll explain the quotation marks below) implies that everything in the Universe, including life, is part of a vast cosmic war between the forces of good (information) and evil (entropy), with entropy the predestined winner. To support this disturbing hypothesis, he weaves a wide-angled interdisciplinary narrative with chapters on cryptography, thermodynamics, classical information theory, genetics, relativity,

quantum mechanics, quantum information theory and cosmology (puzzlingly, string theory is absent). Each of these topics is a major scientific discipline, and Seife seems to have mastered them all.

Taking centre stage throughout the book is Claude Elwood Shannon (1916–2001). Shannon was a brilliant US mathematician–engineer who worked at Bell Labs and in 1948 invented a set of mathematical tools to solve what he called the fundamental problem of communication: “that of reproducing at one point either exactly or approximately a message selected at another point”. Seife omits any



Star wars? The battle between ‘good’ and ‘evil’ in the Universe is really between information and entropy.

LUCASFILM/20TH CENTURY FOX/KOBAL COLLECTION

mention of Shannon's most celebrated result, the 'noisy-channel coding theorem', emphasizing instead Shannon's notion of entropy, the mathematical measure of information and, paradoxically, disorder. Shannon's initial work was indeed called 'information theory', but engineers, mathematicians and physicists today use the term to mean quite different things. I am an information theorist—communication engineer by trade, but Seife's information theory is a branch of physics, and much of the 'information theory' he discusses is very far from communication engineering.

The text is filled with interesting and quirky personalities. In the early chapters, we read of two suicides (Ludwig Boltzmann and Alan Turing) and a guillotining (Antoine-Laurent Lavoisier), but then suddenly all the biographical details stop. I would have liked to see a bit more about the human side of celebrated

intellectuals such as Einstein, Richard Feynman, Stephen Hawking and Shannon, but never mind. According to Seife, in the cosmic war between information and entropy, the individual doesn't matter.

The technical and psychological weight of the material is lightened throughout by the author's unfailing sense of humour, and as homework I assign the reader to locate, somewhere in the text: a scatological remark about McDonalds; a gratuitous wisecrack about the animal-rights group PETA; a discussion about the wisdom of using the Starship Enterprise's transporter beam; and a discussion of the difficulty of getting a cat into a bra.

Finally, potential readers should be warned that this is a non-technical book about some very, very technical material. There is only one equation in the text (apart from footnotes and appendices), which makes it accessible to a

non-mathematical audience but sometimes forces Seife into intricate verbal gymnastics and paradoxes. Indeed, at times Seife seems to doubt his own material, and the book is rife with unscientific adjectives such as absurd, bizarre, confusing, hairy, horrible, insane, ridiculous, sloppy, spooky and just plain weird. (To be fair, the term 'spooky' is bone fide quantum-mechanical technical jargon, as in 'spooky action at a distance'.) By the last chapter, Seife has begun referring to celestial coin flips, god-like beings and supernatural creatures (there are no atheists in black holes, it seems). Still, it will be a rare reader, professional scientist or not, who fails to be entertained and informed by this well-written grand-daddy of all ghost stories. ■

Robert McEliece is in the Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125, USA.

Form becomes feeling

Siobhan Davies looks to science to shape her dance.

Martin Kemp

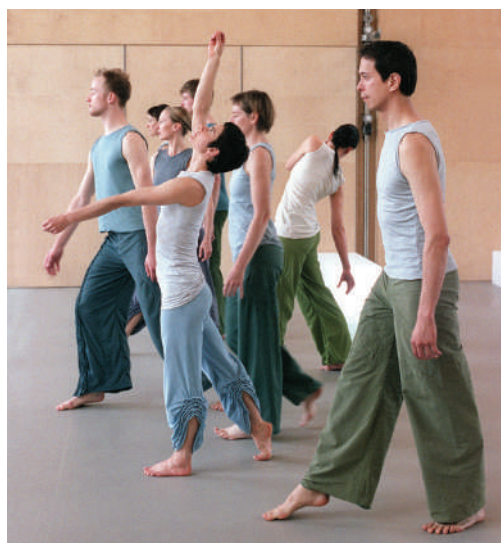
Why does Siobhan Davies expose her dancers to an informal academy of thinkers from diverse worlds, including science and medicine? The answer lies in the depth and power of our instinctive response to body language — a response that choreographers and dancers are adept at exploiting — and her desire to reconstitute and refresh its stock vocabulary.

Her idea is to unsettle the automatic habits and patterns of our responses, especially those involved with dance, refreshing our perceptions by refreshing theirs. The wonder of the complexity of familiar acts, something we take for granted, emerges anew.

For her current programme, *In Plain Clothes*, performed recently in her own stunning dance space and now touring England, she drew in a set of luminaries for her academy. Most notable are the architect Sarah Wigglesworth, the heart surgeon Francis Wells (see *Nature* **429**, 19; 2004), the linguist Susan Hitch and the landscape designer Dan Pearson.

Davies and her dancers witnessed one of Wells's operations. There was no set agenda and no set outcome, other than to promote new shared and individual insights into the human body. The dancers, who are immersed daily in the physical process of dance as a kind of athletic performance, were jolted into a world of new perceptions and thoughts.

The dancers were impressed by the profound stillness of the limbs, torso and head of the anaesthetized patient during the



operation, while the surgeon and his team acted out their own choreographed moves. A heart operation is a highly technical performance, like a dance, but it also has deep emotional resonance for those who bear witness to a life saved. It serves as an analogy for the relationship of form to emotion in dance.

There is also an element of unpredictability in every operation. Davies' work thrives on the confrontation between control and the unpredictable, between imposed order and the unexpected.

In Plain Clothes — the title conveys the direct and unadorned nature of the dance, costumes and setting — began with the structured score of Italian folk songs in a composition by Matteo Fargion. Each dancer was given one of the tunes and asked

separately to assign a movement to each note. This surprisingly literal approach is intended, Davies declares, "to cut out all the fat". It also introduces a random element that sets a challenge for the subsequent ordering of the dancers as an ensemble.

What follows is a phase in which disorder and order fight for formal dominance. There is a process akin to the self-organization of a complex physical process as it settles into a pattern, with the important qualification that intentional choices here guide the emergence of certain orders.

The process is in essence highly formal and abstract. The final result is highly structured, centring on a line of dancers walking laterally across the performance space, casting off and re-absorbing pairs and larger numbers of virtuoso dancers as the piece progresses.

Structure "releases invention", Davies says. Here it creates trigger points at which emotion is released. As in a sonnet, the four-bar blues and folk songs, it is the tightness of the form that allows our perception of the exceptional. Body language works like this in everyday life, with the unexpected — something that departs from the anticipated pattern — jolting us into another level of awareness.

The songs in the piece are about memory, the familiar, departure and loss. Davies does not literally tell their stories. But she uses form as a vehicle to convey the deep elements of their emotional charge.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.

L LEWIS

NEWS & VIEWS



Contrasts — the New York Stock Exchange, symbol of wealth in the United States, and poverty in Haiti.



H. R. ABRAMS/AP

SOCIAL SCIENCES

A New World of differences

Shaun Miller and Jared Diamond

For richer, for poorer — the countries of the Americas and those of the Caribbean present stark contrasts in fortune. An explanation of those contrasts invokes branching chains of cause and effect.

Both the countries of North America (the United States and Canada) are twice as rich as the next richest New World country (Barbados), and some 20 times richer than the poorest (Haiti). The actual numbers are US\$37,562, \$30,677, \$15,720 and \$1,742, respectively¹, expressed here as annual per capita gross domestic product (GDP) corrected for differences in purchasing power. Why are these differences so vast, and can an understanding of the causes help the plight of the poor New World countries? The answers are much debated by social scientists and policy-makers.

One may object: doesn't happiness count more than money? Yes, maybe, but money can buy some prerequisites for happiness. A human development index (HDI) that also considers life expectancy, literacy and school enrolment is moderately correlated with per capita GDP (explained variance 0.51)¹. But intriguing outliers from that correlation reflect countries' differing choices about how to spend their money: Cuba, which is poor, has a deviantly high HDI, whereas Haiti (poor) and the United States (rich) have deviantly low HDIs.

A first step in the analysis is to note that the southernmost countries of the New World —

Argentina and Chile — lie in the temperate zone as do the United States and Canada, and are the richest in Central or South America (GDP \$12,106 and \$10,274, respectively)¹. But both are still three times poorer than the United States or Canada. This suggests that geographical factors furnish part, but not all, of the ultimate answer.

Among the six leading geographical factors, three are linked to a tropical versus a temperate location. First, tropical diseases such as malaria and yellow fever formerly caused high mortality and hence low immigration of European settlers², and still depress economic productivity today³. Second, agricultural productivity is lower in tropical than in temperate climates: for instance, although maize originated in the tropics, its output per hectare today in the five countries lying beyond latitude 30° N or 30° S (the United States, Canada, Argentina, Chile and Uruguay) is three to nine times the value for any tropical New World country⁴. Third, before the advent of air-conditioning, machinery in tropical factories had more frequent breakdowns, and it now incurs higher air-conditioning costs⁵. A fourth geographical factor — land-locked location —

penalizes Bolivia and Paraguay, the only countries in the New World that do not have their own ocean ports, or rivers deep enough for ocean-going ships, because transportation costs overland are seven times those by sea⁶.

The final two geographical factors, Native American population density and mineral abundance, affect prosperity partly through their effects on human institutions. Europeans created three types of New World colony, each differing in its population make-up and economy, and hence in its institutions, whose legacies still have heavy consequences today^{2,7,8}. In one type of colony (for example, Bolivia, Peru and Mexico), dense native populations were exploited as labour to extract rich mineral resources. In a second type (for example, Brazil, Barbados, Cuba, Jamaica and Haiti), imported slaves were used to work large tropical plantations that grew high-value crops (especially sugar and coffee). Both of those types of colony were ruled by small European élites (less than 25% of the population), who institutionalized political inequality to exploit native or slave labour.

In contrast, in the third type of colony ('neo-Europes'), few mineral resources were

P. TURNLEY/CORBIS

exploited in colonial times; the native populations were sparser (at least after the impact of European diseases); and their temperate locations ruled out tropical plantations but permitted the dense settlement of Europeans, who came to comprise 85–97% of the population. (The southern United States did have slave plantations, but they were smaller and less profitable than those in tropical colonies.) Most of the European settlers became small farmers in communities that lacked a dominant élite and were unable to extract wealth from natives or slaves. The neo-Europes were the United States and Canada, and to some extent their temperate South American counterparts (Argentina, Chile and Uruguay)^{2,7,8}.

As a result, the New World has undergone a reversal of fortune, by which the areas poorest at the time of European arrival (temperate North and South America) are the richest today, whereas the richest areas in Native American times (Bolivia, Peru and Mexico) are poor today^{2,7,8}. Haiti, today the New World's poorest country, was the richest in 1790, while Voltaire was describing the former poverty of the modern United States and Canada when he wrote that the British and French at war to control North America in the 1750s were "fighting over a few acres of snow".

This reversal of fortune reflects the persistent modern obstacles to investment and development that were created by extractive institutions set up by European élites in the formerly rich colonies. And much of the population of such colonies today is poorly educated and contributes little to the economy. The reversal applies not only to the New World as a whole, but also to some individual regions that in essence constitute natural experiments. For instance, among the five adjacent Central American nations originally united under Spanish colonial rule, Costa Rica was formerly the poorest but is now 2.5 times richer than the other four (Guatemala, El Salvador, Honduras and Nicaragua)¹.

Natural resources

Another paradoxical effect of geography is the "curse of natural resources"^{9,10}. It is naively expected that natural resources would tend to make a nation rich. In fact, the oil and mineral booms that Bolivia, Peru, Mexico and Venezuela experienced from 1970 to 1990 contributed to a drop rather than a rise in GDP, by competing with other economic sectors, fostering inequality and corruption, and discouraging long-term economic development and spending on education.

These are some of geography's effects on New World prosperity and happiness. But institutions have also diverged in the New World for historical reasons, especially whether an area became a Spanish, Portuguese or British colony. These three mother countries differed in their own institutions, which became transferred to or influenced the independent nations later emerging from the

"In the twentieth century there were no military coups in the United States or Canada, but an average of 8.5 coups in almost all Central and South American countries."

colonies¹¹. Spain and Portugal had stronger kings, less democratic governments and more institutionalized inequality than did Britain. British society, more than Spanish or Portuguese society, shared beliefs in devolved forms of government — reinforced in the United States by immigrants seeking religious freedom, and in Canada and the United States by the democratizing effects of the frontier. Spain further enforced central control by restricting its New World colonies' trade with Spain to just six ports, and by requiring trade between its colonies to be routed via Spain. In contrast, Britain let its colonies trade with each other and with Britain via hundreds of ports¹².

The result in former Spanish and Portuguese colonies was chronic fighting between each national government and local groups, and between the élite and non-élite, for power¹². For instance, in the twentieth century there were no military coups in the United States or Canada, but an average of 8.5 coups in almost all Central and South American countries¹³. Government expropriation of private production discouraged investment and increased inequality in Spanish and Portuguese colonies. Even today, the so-called GINI index (a measure of economic inequality) is lower, and the richest 10% of the population enjoys relatively less income and consumption, in the United States and Canada than in any Central or South American country¹⁴.

The Industrial Revolution began in Britain and spread easily to North America because of trade and the common language of English³. But it began in Spain and Portugal only in the late nineteenth century, and it was opposed by their former colonies' élites because those countries had poor protection of private property rights, and their élites would benefit little and even lose political power through industrialization⁸. For instance, by 1900 both the United States and Canada had a length of railroad track (relative to population or to the country's length) several times that of any other New World country, except for Argentina⁴.

Many of the factors outlined above converge on the proximate explanation for wealth that is preferred by many economists: institutions that reward private investment. One such institution is legal protection of private property rights, which is important for the owner's ability to mortgage and sell property, attract investment and do business¹⁵. If a house can be mortgaged, it becomes not just a place in which to sleep but also a means to create

investment capital. For instance, most homes in the United States are mortgaged, whereas most property in developing countries is not, and most of those US mortgages are used to fund new businesses. Protection of private property rights can be variously measured by safeguards against appropriation, domestic credit provided by banks, and the percentage of a country's adults who are recorded as borrowers. All three measures are higher for the United States and Canada than for any other New World country, except Costa Rica¹⁴.

Economic examples

To some readers, this long list of reasons behind North America's prosperity and Central and South America's comparative poverty will seem depressing. Others will see it as the basis of recipes for attacking poverty. Economic examples abound. Legal protection of property rights in developing countries really does let property owners become investors. Costa Rica ended the possibility of military coups by abolishing its army in 1949, and it is now the second richest country in Central America. Trinidad and Tobago is now the fifth richest New World country south of the United States, despite a resource boom that has made it the Caribbean's biggest producer of oil and natural gas. It has escaped the curse of natural resources through high expenditure on education, low corruption, economic diversification and sound monetary policy. Thus, government, history and today's choices all make a difference. ■

Shaun Miller and Jared Diamond are in the Department of Geography, University of California, Los Angeles, California 90095-1524, USA.
e-mails: jdiamond@geog.ucla.edu; smill24@gmail.com

1. United Nations Human Development Report 2005: *International Cooperation at a Crossroads: Aid, Trade and Security in an Unequal World* (UN Development Programme, New York, 2005).
2. Acemoglu, D., Johnson, S. & Robinson, J. A. *Am. Econ. Rev.* **91**, 1369–1401 (2001).
3. Sachs, J. D. *The End of Poverty: Economic Possibilities for Our Time* (Penguin, New York, 2005).
4. Mitchell, B. R. *International Historical Statistics: The Americas 1750–2000* 5th edn (Palgrave Macmillan, New York, 2003).
5. Blum, B. S. *Cuadernos de Economía* **40**, 423–433 (2003).
6. Limao, N. & Venables, A. J. *World Bank Econ. Rev.* **15**, 451–479 (2001).
7. Engerman, S. L. & Sokoloff, K. L. in *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800–1914* (ed. Haber, S.) 260–304 (Stanford Univ. Press, 1997).
8. Acemoglu, D., Johnson S. & Robinson, J. A. *Q. J. Econ.* **117**, 1231–1294 (2002).
9. Sachs, J. D. & Warner, A. M. *J. Dev. Econ.* **59**, 43–76 (1999).
10. Ross, M. L. *J. Peace Res.* **41**, 337–356 (2004).
11. North, D. C. *Understanding the Process of Economic Change* (Princeton Univ. Press, 2005).
12. North, D. C., Summerhill, W. & Weingast, B. R. in *Governing for Prosperity* (eds de Mesquita, B. B. & Root, H. L.) 17–58 (Yale Univ. Press, 2000).
13. Smith, P. H. *Democracy in Latin America: Change in Comparative Perspective* (Oxford Univ. Press, 2005).
14. World Bank 2005 *World Development Indicators* (World Bank, Washington DC, 2005).
15. De Soto, H. *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else* (Basic, New York, 2000).

GENETICS

Paramutable possibilities

Paul D. Soloway

A curious genetic phenomenon allows certain genetic instructions to be passed between generations without the gene variants involved being transmitted. Some spotty mice provide clues to how this might happen.

One of the benefits of sex is that our parents give us two copies of nearly all our genes. So if one copy, or allele, doesn't work, there is a good chance that the other, operating independently, will be sufficient. Exceptions exist, but even in those cases the two alleles usually function independently. In rare cases, however, the alleles interact intimately, even though they reside on separate chromosomes, and whether one allele is expressed depends on directions given by the other. One example of this phenomenon is paramutation, a process in which orders issued by an allele in one generation are remembered in subsequent generations, even if the allele issuing the order is not transmitted. How do chromosomes remember who their partners were in previous generations, and how do those partners continue to exert their control long after they have segregated away? On page 469 of this issue, Minoou Rassoulzadegan *et al.*¹ provide evidence for a mechanism involving small RNA molecules that might underlie paramutation.

In 1956, R. Alexander Brink² coined the term paramutation to describe the unexpected means of regulating purple pigmentation in maize kernels (see ref. 3 for a review). Pigmentation is controlled, in part, by the *R* gene locus. If all the alleles at this locus are recessive alleles (r'), the kernels are colourless, but when a single dominant allele (R^R) is present, the kernels are typically purple. This is not always the case, however; it depends on what the R^R allele was exposed to in the previous generation. In Brink's experiments, if the R^R allele came from pollen of a plant that also had the R^{st} allele, which caused the kernels to be spotted, then the normally dominant influence of the R^R allele was silenced and kernels showed reduced pigmentation. The untransmitted R^{st} allele exerted its effects in the next generation, overriding the dominance of the R^R allele in violation of Mendel's rules of genetic inheritance. Brink called R^R paramutable and R^{st} paramutagenic, and he showed that the influence of the paramutagenic allele could persist for many generations. Paramutation-like phenomena have been reported in mammals^{1,4-6} and might be of clinical relevance in humans, for example affecting certain forms of diabetes⁴.

The system examined by Rassoulzadegan *et al.* in mice involves an allele of the *Kit* locus called *Kit^{tm1Alf}*. This is a null allele; that is, it makes no functional Kit protein. Heterozygous mice — those with one *Kit^{tm1Alf}* allele and

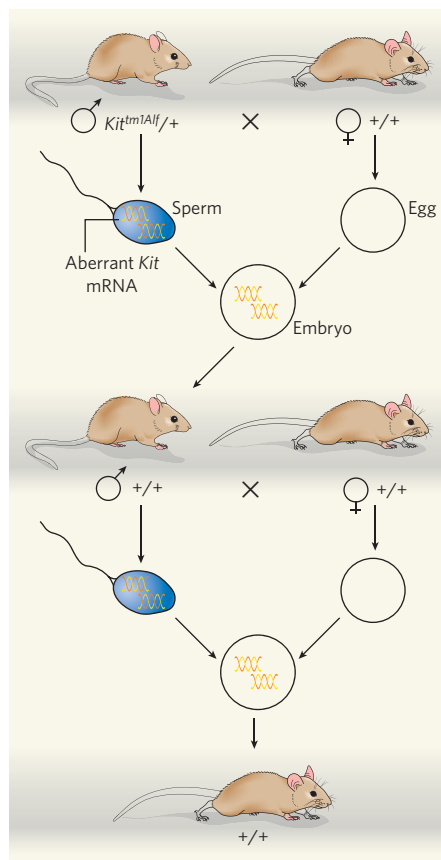


Figure 1 | Model for paramutation at *Kit* as proposed by Rassoulzadegan *et al.*¹ Heterozygous male mice with one *Kit^{tm1Alf}* allele and one normal wild-type (+) allele have a spotted white tail-tip and produce aberrant *Kit* messenger RNAs from the paramutagenic *Kit^{tm1Alf}* allele. These are packaged in sperm and transmitted to the embryo after fertilization of the egg. Progeny carry a wild-type *Kit* allele transmitted by the heterozygous father, but action of the transmitted aberrant RNAs still gives rise to the spotted tail and maintains their production, allowing paramutation of the wild-type *Kit* allele and further transmission of the spotted tail. Loss of aberrant RNAs through attenuated production or dilution over successive generations might lead to gradual loss of paramutation.

one normal 'wild-type' *Kit* allele — had white spotting at their tail tips and showed reduced expression of *Kit* (as measured by production of *Kit* messenger RNA from the gene). When these mice were crossed with wild-type mates, the wild-type progeny showed the same white spotting and reduced *Kit* mRNA levels as their heterozygous parent, even though they were

fully wild type and lacked the null allele that caused spotting in their heterozygous parent. Moreover, these wild-type progeny also had an accumulation of a mixture of smaller RNAs with sequences that matched various parts of the *Kit* mRNA. Notably, sperm precursor cells from male mutants had aberrant *Kit* mRNAs, and their sperm showed an unexpected accumulation of RNA. It is possible that these RNAs are transmitted to the next generation on fertilization, even if the allele from which they arose is not passed on.

The nature of aberrant RNAs is unknown, but they have attracted attention for a variety of reasons. First, paramutation is one of several heritable 'epigenetic' phenomena, in which inherited trait differences arise by covalent modifications to DNA and DNA-bound histone proteins rather than by changes to the DNA sequence of the genes themselves. These modifications can be regulated by small interfering RNAs (siRNAs) generated by the RNA interference (RNAi) pathway, and can arise from transcribed DNA repeats⁷⁻⁹, which are structures found at some loci undergoing paramutation^{4,6,10}. Second, siRNAs and the related microRNAs (miRNAs) can cause the degradation of mRNAs whose sequences they share, blocking production of the encoded proteins¹¹. Finally, RNAs have been proposed to play a role in another non-mendelian system of inheritance¹². In combination with Rassoulzadegan and colleagues' data, these observations raise the possibility that the aberrant RNAs arising from the *Kit^{tm1Alf}* allele include siRNAs, miRNAs or other regulatory RNAs that are packaged in sperm and cause paramutation on transmission to the next generation.

To test this possibility, Rassoulzadegan *et al.* injected total RNA from tissues containing the aberrant *Kit* RNAs into fertilized mouse eggs. This led to spotting in many of the progeny that came to term — and in their progeny. Injected miRNAs designed to degrade *Kit* mRNA had the same effect. Control miRNAs with no similarity to *Kit* unexpectedly also caused spotting, but at a lower frequency, and the spotted animals only rarely transmitted the spotted phenotype to their progeny. It is not certain that spotting caused by the *Kit^{tm1Alf}* allele arises by the same mechanism as spotting induced by injected miRNAs. However, what might be occurring in this system, and possibly in other paramutation models, is that small RNAs produced from a paramutagenic allele are acting on the corresponding paramutable allele or on its transcribed mRNA, effectively silencing it. Because RNAi-mediated degradation of mRNAs produces more siRNAs, the silencing might be propagated if these small RNAs are packaged into germ cells and carried into the next generation. This could allow successive generations to display a certain characteristic, even if the paramutagenic allele that caused it was not transmitted (Fig. 1). Rassoulzadegan and colleagues' proposal that RNAs are involved

in paramutation is strongly supported by work from Vicki Chandler's group showing that paramutation requires an RNA metabolizing enzyme that is involved in other epigenetic phenomena¹³.

Rassoulzadegan and colleagues' model has yet to be validated, and several points need to be clarified. Notably, the mechanism by which aberrant RNAs from *Kit^{tm1Alf}* heterozygotes induce spotting is not known. Does it involve bona fide siRNAs or miRNAs emanating from the mutant allele? Also, are aberrant RNAs causing mRNA degradation or epigenetic modifications at the wild-type allele? Are these effects mediated by RNAi or by other pathways influenced by small RNAs? If siRNAs or miRNAs do emanate from the paramutagenic *Kit^{tm1Alf}* allele, how do they arise, and do similar RNAs arise in other paramutation models? A particularly intriguing possibility is that such RNAs regulate other non-genetic modes of inheritance, such as metabolic or behavioural imprinting. These have far greater consequences for human development than for

spotty mice and maize, but we may learn about such mysterious processes by studying those mouse tales.

Paul D. Soloway is in the Division of Nutritional Sciences, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York 14853, USA.
e-mail: pds28@cornell.edu

1. Rassoulzadegan, M. *et al.* *Nature* **431**, 469–474 (2006).
2. Brink, R. A. *Genetics* **41**, 872–889 (1956).
3. Chandler, V. L. & Stam, M. *Nature Rev. Genet.* **5**, 532–544 (2004).
4. Bennett, S. T. *et al.* *Nature Genet.* **17**, 350–352 (1997).
5. Rassoulzadegan, M., Magliano, M. & Cuzin, F. *EMBO J.* **21**, 440–450 (2002).
6. Herman, H. *et al.* *Nature Genet.* **34**, 199–202 (2003).
7. Aufsatz, W., Mette, M. F., van der Winden, J., Matzke, A. J. & Matzke, M. *Proc. Natl Acad. Sci. USA* **99**, 16499–16506 (2002).
8. Hamilton, A., Voinnet, O., Chappell, L. & Baulcombe, D. *EMBO J.* **21**, 4671–4679 (2002).
9. Volpe, T. A. *et al.* *Science* **297**, 1833–1837 (2002).
10. Stam, M., Belele, C., Dorweiler, J. E. & Chandler, V. L. *Genes Dev.* **16**, 1906–1918 (2002).
11. Fire, A. *et al.* *Nature* **391**, 806–811 (1998).
12. Lolle, S. J., Victor, J. L., Young, J. M. & Pruitt, R. E. *Nature* **434**, 505–509 (2005).
13. Alleman, M. *et al.* *Nature* (in the press).

NANOMATERIALS

Display of flexibility

László Forró

Treated the right way, carbon nanotubes can be moulded into large, flexible electron-emitting sheets. The material is one half of what's needed for an electronic display you could fold up and slip in your pocket.

As far as their electrical properties are concerned, the long, cylindrical carbon molecules known as carbon nanotubes are notoriously Janus-faced: picking out a nanotube from a newly fabricated batch, one cannot know whether it will be a conductor or an insulator. In principle, both types are useful, and the basic transport, mechanical and optical properties of each kind are more or less known. But not knowing how to tune the synthesis to yield just one type is a considerable handicap to the use of nanotubes as, for example, components in integrated circuits. Because of this, attention has concentrated on applications for which this uncertainty is irrelevant: nanoswitches¹, motors², actuators³ and yarns for composite materials⁴, to name but a few.

One particularly promising application of carbon nanotubes is the flat-panel display. Such displays exploit the fact that, owing to quantum-mechanical tunnelling effects, nanotubes are very efficient emitters of electrons when placed under electric fields. They thus act as a source of electrons that is similar to the cathode-ray tube used in conventional monitors, but one that is just a few millimetres thick and operates at a fraction of the power. The Samsung Advanced Institute of Technology in South Korea has recently reported the

fabrication of a 30-inch flat television screen, based on field emission from carbon nanotubes, that is close to commercialization⁵. Now, writing in *Nano Letters*, Yung Joon Jung and colleagues⁶ report the assembly of the field-emission part of a screen using carbon-nanotube electrodes embedded in a polymer matrix. Their simple idea could be a big step towards the implementation of large-area displays that are not only flat, but flexible too.

The authors' method consists of patterning islands of catalytic particles on a surface of silicon dioxide. Using the technique of chemical vapour deposition, they synthesize vertically aligned carbon-nanotube pillars 500 micrometres in diameter and 100 micrometres high. This large area of nanotube pillars is impregnated with dimethylsiloxane, a polymer precursor. Heating the resulting matrix to 100 °C polymerizes it, and the nanotube-polymer composite can be peeled off (Fig. 1). These embedded nanotube pillars show excellent emission characteristics, with a field-enhancement factor (defined as the ratio of the electric field at the tip of the pillar to the applied electric field) of 10,000.

The originality of the work lies in the polymer precursor, which perfectly wets the individual nanotubes within the pillars. Without



Figure 1 | Flexible friend. Jung and colleagues⁶ nanotube-polymer composite, with its evenly spaced carbon pillars clearly visible.

this wetting, surface tension would have shrunk the pillars, and the mutual screening of the electric field at tips of neighbouring nanotubes would have damped down the exceptional field-enhancement factor. Crucially, the nanotubes preserve their emission properties even if the matrix is severely bent.

Although this highly flexible nanotube-polymer composite could prove to be a central part of a future foldable flat-screen display, there is still a long way to go. A second, essential component will be a flexible screen that, as in a conventional display, fluoresces when struck by electrons sent out by the field-emitting part. This screen must be kept at a constant separation from the field-emitting composite, even when the display is bent. Constructing such a component will be difficult, but is not an insurmountable task.

The potential applications of flexible displays are many. One could conceive of a flexible electronic newspaper, the pages of which are reloaded using wireless network technology, and which you could bend or roll up after reading. More whimsically, projecting a car's surroundings onto its body with such a display would make it invisible — allowing James Bond to die another day. Perhaps one could even envisage a time when the opera diva changes her dress between scenes by simple reprogramming.

Composites of carbon nanotubes and polymers are not the only candidates to inspire such unbridled imagination. Organic light-emitting diodes are another, but these have their own problems, suffering from short lifetimes owing to air sensitivity, fatigue and the like. Thus, contributions such as that of Jung and colleagues⁶ play an important part in ensuring the emergence of the flexible display from the realm of science fiction to that of science fact.

László Forró is at the Institute of Physics of Complex Matter, FBS Station 3, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.
e-mail: laszlo.forro@epfl.ch

- Cummings, J., Collins, P. G. & Zettl, A. *Nature* **406**, 586 (2000).
- Fennimore, A. M. *et al.* *Nature* **424**, 408–410 (2003).
- Ke, C. & Espinosa, H. D. *Appl. Phys. Lett.* **85**, 681–683 (2004).
- Zhang, M. *et al.* *Science* **309**, 1215–1219 (2005).
- Han, I. T. *et al.* invited talk at XIX Winterschool/Euroconference on the Electronic Properties of Novel Materials, Kirchberg, Austria (2005).
- Jung, Y. J. *et al.* *Nano Lett.* **6**, 413–418 (2006).

IMMUNOLOGY

Adaptable innate killers

Peter Parham

Natural killer cells are versatile white blood cells that act in the innate immune system. Quite how adaptable they can be in the absence of other, more specialized, immune cells comes as a surprise.

Contact dermatitis is an occupational hazard for many people. This disruptive immune reaction can be caused by contact with an odd assortment of items, including plants, drugs, chemicals and even money¹. In each case, small, reactive molecules from the item penetrate the skin and modify proteins within the tissue. Thus altered, the skin proteins are seen as foreign by the immune system, which then endeavours to destroy them. This dermatitis is caused by an adaptive response²: the reaction occurs only if that specific antigen has been encountered before, and it is mediated by the archetypal white blood cells, or lymphocytes, of adaptive immunity — activated antigen-specific $\alpha\beta$ T cells. In contact sensitivity, the mouse experimental model of contact dermatitis, three other types of lymphocyte are implicated in the reaction: B-1 B cells, NK T cells and $\gamma\delta$ T cells³. These cells are part of the fast-acting defences of innate immunity, which do not strengthen with successive exposures to the antigen. Notably absent from this list is the natural killer cell. So the report by O'Leary *et al.* in *Nature Immunology*³ is unexpected. They find that in mice lacking all other lymphocytes, natural killer cells play all the roles necessary to produce contact sensitivity.

Natural killer cells are different from other lymphocytes because during cell maturation they do not need to rearrange the genes for their antigen receptors (T-cell receptors or immunoglobulins). In mutant mice lacking the *Rag2* gene (*Rag2*[−] mice), such gene rearrangement is blocked, so B cells, T cells and NK T cells do not exist, and natural killer cells are the only circulating lymphocytes. Unexpectedly, contact-sensitivity reactions against the chemical dinitrofluorobenzene (DNFB) (Fig. 1) are almost as strong in *Rag2*[−] mice as in normal mice³. In exploring this observation, O'Leary *et al.* show that, in the absence of other lymphocytes, natural killer cells are necessary for the contact-sensitivity reaction. They also find the reaction to be antigen-specific, so that *Rag2*[−] mice primed with DNFB respond to further contact with DNFB but not to a first

contact with a different reaction chemical.

Not too long ago, natural killer cells were considered to be all exactly the same and lacking any antigen specificity. But O'Leary *et al.* add to a body of evidence showing how specific subpopulations of natural killer cells are activated in response to bone-marrow transplants^{4,5}, viral infections⁶, tumours⁷ and, now, as part of contact-sensitivity reactions to chemically modified skin tissue³. Natural-killer-cell subpopulations are distinguished by their differential expression of a variety of activating and inhibitory receptors, of which the Ly49 receptors are notably diverse and variable among individuals⁸. Contact sensitivity to DNFB is mediated by the subpopulation of natural killer cells expressing Thy1, an activation marker, and either the Ly49C or Ly49I receptor. These cells concentrate in the liver³, where DNFB is probably detoxified.

The experiments of O'Leary *et al.* touch upon immunological memory, the property that has most clearly distinguished innate from adaptive immunity. Surviving a childhood infection with mumps or measles virus provides immunity from the disease for the rest of a person's life. In current thinking, this protection is provided by 'memory' B and T cells, as part of adaptive immunity, that emerge during the immune response to the infection and thereafter ensure a quick, strong and overwhelming response to further infection.

After priming *Rag2*[−] mice with DNFB, O'Leary *et al.* show that a secondary challenge, made either 5 or 28 days later, gave a strong response (Fig. 1). Their interpretation, that priming resulted in a persistent memory mediated by natural killer cells, is provocative and should stimulate investigation. For example, the difference between 5 and 28 days is not a particularly demanding test of persistent memory, even for a laboratory mouse with a lifespan of around two years. And, in O'Leary and colleagues' experimental design, it is possible that the primary response to DNFB had not truly subsided by the time of the secondary challenge. Elimination of DNFB and of DNFB-mediated tissue components from the

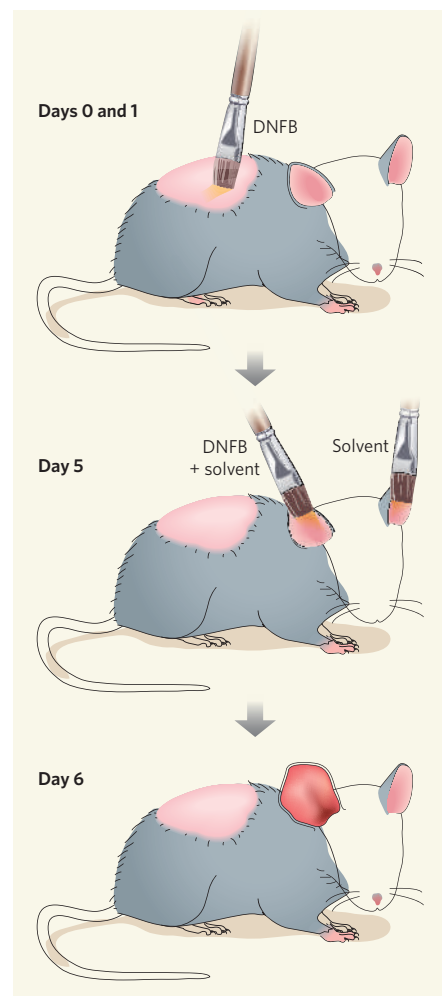


Figure 1 | Measuring contact sensitivity. The shaved back of a mouse is painted with a reactive chemical, such as dinitrofluorobenzene (DNFB). This first contact is the priming stage, in which the adaptive immune response to DNFB-modified tissue is induced. The second contact is made five days later, when one ear of the mouse is painted with DNFB and the other, the control, is painted with the solvent in which the DNFB is dissolved. One day after this secondary challenge the immunological reaction is measured by comparing the thickness of the two ears, as the ear painted with DNFB is enlarged by inflammation caused by infiltrating immune cells and plasma. O'Leary *et al.*³ found that the strength of the response to DNFB in *Rag2*[−] mice, which have natural killer cells but no other lymphocytes, was comparable in strength and specificity to that in normal mice.

mouse's painted back could take considerable time, resulting in persistent stimulation, rather than the development of memory, throughout the experiment's duration. This sort of persistent stimulation is what causes occupational contact dermatitis in humans, for instance when sensitized supermarket clerks daily handle nickel-containing coins.

In the course of evolution, lymphocytes became more diverse and more specialized. But dependency accompanies specialization, as witnessed in the *Rag2*[−] mouse — B cells, NK T cells and T cells are so reliant on their

particular antigen receptors that they die when they cannot make them. By contrast, natural killer cells use many different receptors and are less dependent upon any one of them. The Ly49 receptors are a case in point, as they are flourishing in mice but do not occur in human natural killer cells⁹.

The study of O'Leary *et al.* vividly illustrates the functional versatility and adaptability of natural killer cells. In normal mice, where contact sensitivity has little natural-killer-cell contribution, the response involves at least four other lymphocyte types. In *Rag2*[−] mice, in which natural killer cells flourish in the absence of other lymphocytes, they are able to produce an equivalent contact sensitivity by themselves. A parallel can be seen in bone-marrow transplantation, where eliminating T cells from the graft allows helpful natural killer cells to emerge⁵. In their adaptability,

versatility and deference to more specialized lymphocytes, natural killer cells seem to be the modern lymphocyte most like those lymphocyte ancestors that populated immune systems before the evolution of antigen-receptor genes that required rearrangement. ■

Peter Parham is in the Department of Structural Biology, Stanford University, California 94305-5126, USA.
e-mail: peropa@stanford.edu

1. Saary, J. *et al.* *J. Am. Acad. Dermatol.* **53**, 845 (2005).
2. Askenase, P. W. *Clin. Exp. Immunol.* **125**, 345–350 (2001).
3. O'Leary, J. G., Goodarzi, M., Drayton, D. L. & von Adrian, U. H. *Nature Immunol.* **7**, 507–516 (2006).
4. George, T. *et al.* *Immunol. Rev.* **155**, 29–40 (1997).
5. Ruggeri, L. *et al.* *Science* **295**, 2097–2100 (2002).
6. Arase, H., Mocarski, E. S., Campbell, A. E., Hill, A. B. & Lanier, L. L. *Science* **296**, 1323–1326 (2002).
7. Pende, D. *et al.* *Blood* **105**, 2066–2073 (2005).
8. Makrigrannis, A. P., Patel, D., Goulet, M. L., Dewar, K. & Anderson, S. K. *Genes Immun.* **6**, 71–83 (2005).
9. Parham, P. *Nature Rev. Immunol.* **5**, 201–214 (2005).

METEORITICS

How to make a chondrule

Steve Desch

Chondrules, the stony, seed-like grains in meteorites, were formed when some event melted rock in the solar nebula. The latest analyses narrow the possible 'when', 'where' and 'how' of that process.

Open up almost any stony meteorite, as scientists have been doing for more than 200 years¹, and you will find hundreds of millimetre-sized bits of rock. These 'chondrules' (named after the Greek for seeds) were formed at the birth of the Solar System, and as such potentially bear witness to conditions — pressures, temperatures, chemical composition and so on — in the solar nebula. But obtaining that information depends on identifying how they formed, a topic tackled by Cuzzi and Alexander on page 483 of this issue².

With the invention of the petrographic microscope in the late nineteenth century, it was recognized that chondrules are silicate rocks with igneous textures, and that these "drops of fiery rain"³ individually crystallized from a molten state while floating freely. Clearly, at the birth of the Solar System there were events hot enough to melt rock, and frequent enough to melt most of the mass of the asteroids — asteroids, which reside between Mars and Jupiter today, are the source of most meteorites. Identifying what melted the chondrules has long been a central theme of meteoritics. Dozens of explanations have been advanced, but detailed, quantitative tests of these models have come of age only recently.

Cuzzi and Alexander's work² represents a major advance in constraining the circumstances of chondrule formation. It shows that not only were the chondrule-melting events very energetic, but they were also very large —

thousands of kilometres — in extent. This finding limits the number of possible mechanisms by which chondrules formed, and goes some way to demystifying their origins.

Chondrules are known to have reached peak temperatures in excess of 1,800 K, and to have remained partially molten for a matter of hours⁴. This degree of heating not only melts the silicate rock, but can also cause relatively volatile elements such as potassium, iron, silicon and magnesium to evaporate. Indeed, these elements are relatively depleted in chondrules compared with the average in meteorites. Because light isotopes evaporate more readily, one would expect the Si in chondrules, for example, to be relatively depleted in light isotopes such as ²⁸Si compared with heavy isotopes such as ³⁰Si. In fact, no such isotopic fractionation is observed⁵.

As Cuzzi and Alexander² discuss, the most plausible explanation for this discrepancy is that the K, Fe, Si and Mg vapour never left the region in which the chondrules formed. The chondrules would then reequilibrate with the vapour of nearby chondrules, and the light isotopes that most readily evaporate from chondrules would just as quickly recondense onto them. Based on the upper limit on the isotopic fractionation, and using an elegantly simple analytic expression confirmed by numerical chemical kinetic modelling, Cuzzi and Alexander estimate that the vapour pressures of the volatile elements must have reached at

least 95% of their saturation pressures for the duration of the chondrule-melting event. This constraint sets a lower limit on the density of chondrules floating in the chondrule-formation region of about 10 per cubic metre.

An additional constraint arises from the need to keep the rock vapour from diffusing away from the chondrules, which sets a lower limit on the size of the chondrule-forming region. In the region where meteorites are thought to have formed (the location of the present-day asteroid belt), gas pressures probably were about 10^{−4} bar (ref. 6). In that case, a chondrule would have to be surrounded by other evaporating chondrules for about 3 km in each direction to keep the rock vapour pressure close to saturation. For the majority (99%) of chondrules to escape isotopic fractionation, the chondrule-forming region would have to be about 300 times larger still — that is, about 1,000 km in extent.

These new estimates severely restrict the 'when', 'where' and 'how' of chondrule melting. Previous studies of how often chondrules collided with and stuck to each other led to estimates of 1–10 chondrules per cubic metre in the chondrule-forming region⁷, and it is significant that these new constraints from isotopic fractionation confirm this high density. Chondrule densities of as much as 10 per cubic metre are unexpectedly high and can exist only where the gas densities are highest, near the midplane of the Solar System's protoplanetary disk. Even so, chondrules must be concentrated by factors of several hundred compared with their average throughout the disk, so that they outweigh the gas locally. This argues against heating mechanisms acting far from the midplane, such as shocks driven by X-ray flares⁸ or clumpy accretion⁹. The constraint on the size of the chondrule-formation region, 1,000 km or more, is fundamentally new: it constitutes evidence against such proposed heating mechanisms as nebular lightning¹⁰, and perhaps even shock waves driven by large asteroids, several hundred kilometres in diameter, orbiting through the solar nebula gas¹¹.

All in all, Cuzzi and Alexander² have notably advanced our understanding of chondrule formation. On the evidence of their analyses, a favoured explanation is that the melting mechanism for chondrule creation was shock waves, on a scale of the whole solar nebula^{12–14}. These shock waves were probably driven by gravitational instabilities that arose when the gravitational forces of the solar nebula gas on itself were comparable to the gravitational pull of the Sun. These instabilities probably manifested themselves as spiral density waves akin to the spiral arms in our Galaxy. Further modelling, in concert with petrological and isotopic analyses of chondrules, will test this idea. Despite their size, smaller than seeds of grain, chondrules may bear witness to events that occurred on a stage as large as the Solar System itself. ■

Steve Desch is in the Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287-1504, USA.
e-mail: steve.desch@asu.edu

- Howard, E. *Phil. Trans. R. Soc. Lond.* **92**, 168–212 (1802).
- Cuzzi, J. N. & Alexander, C. M. O'D. *Nature* **441**, 483–485 (2006).
- Sorby, H. L. *Nature* **15**, 495–498 (1877).
- Hewins, R. H., Connolly, H. C. Jr, Lofgren, G. E. & Libourel, G. in *Chondrites and the Protoplanetary Disk* Vol. 341 (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) 286–316 (Astron. Soc. Pacif. Conf. Ser., Provo, Utah, 2005).
- Davis, A. M., Alexander, C. M. O'D., Nagahara, H. & Richter, F. M. in *Chondrites and the Protoplanetary Disk* Vol. 341 (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) 432–455 (Astron. Soc. Pacif. Conf. Ser., Provo, Utah, 2005).

- Bell, K. R., Cassen, P. M., Klahr, H. H. & Henning, Th. *Astrophys. J.* **486**, 372–387 (1997).
- Gooding, J. L. & Keil, K. *Meteoritics* **16**, 17–43 (1981).
- Nakamoto, T., Hayashi, M. R., Kita, N. T. & Tachibana, S. in *Chondrites and the Protoplanetary Disk* Vol. 341 (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) 883–892 (Astron. Soc. Pacif. Conf. Ser., Provo, Utah, 2005).
- Boss, A. P. & Graham, J. A. *Icarus* **106**, 168–178 (1993).
- Desch, S. J. & Cuzzi, J. N. *Icarus* **143**, 87–105 (2000).
- Ciesla, F. J., Hood, L. L. & Weidenschilling, S. J. *Meteorit. Planet. Sci.* **39**, 1809–1821 (2004).
- Wood, J. A. *Meteoritics* **31**, 641–645 (1996).
- Desch, S. J. & Connolly, H. C. Jr *Meteorit. Planet. Sci.* **37**, 183–207 (2002).
- Desch, S. J., Ciesla, F. J., Hood, L. L. & Nakamoto, T. in *Chondrites and the Protoplanetary Disk* Vol. 341 (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) 849–872 (Astron. Soc. Pacif. Conf. Ser., Provo, Utah, 2005).

ECOLOGY

Paradox of the clumps

Sean Nee and Nick Colegrave

A fresh look at an established model in ecology has generated insights into how species coexist with each other. But it has also raised a vexed question: what constitutes the ecological identity of species?

Classical ecology discovered the principle of competitive exclusion — or, more pithily, 'one species, one niche'. In order to coexist, species must have their own individual way to make a living, otherwise the superior competitors would exclude the inferior. Niches might pre-exist: for example, if there are two types of seed in the environment, this provides two niches for specialist seed-eating birds. Or niches might be created by a species evolving into openings in the 'marketplace' of their ecology.

As they report in *Proceedings of the National Academy of Sciences*, Scheffer and van Nes¹ have revisited a well-studied classical model of competing species and discovered something new. Even in the absence of any environmental discontinuities, they find that assemblages of species will self-organize into clumps of species with very similar niches within a clump and a large difference between clumps. So, paradoxically, species both do, and do not, organize themselves into discrete niches.

In the *Origin of Species*, Darwin asked: "Why, if species have descended from other species by fine gradations, do we not everywhere see innumerable transitional forms? Why is not all nature in confusion, instead of the species being, as we see them, well defined?" This evolutionary question has a closely related ecological counterpart: how similar can species be to one another and still coexist? A well-studied model for this problem is outlined in Figure 1. The question becomes this. With a single, continuous niche axis, how many species can you pack along it? Or, is there a limit to how close the species can be along this axis?

Although some confusion persists about the answer, the canonical result^{2,3} is that there are, indeed, limits to similarity, and coexistence is

possible only for species spaced out along the axis. In other words, we will observe species occupying distinct, spaced niches even in the absence of environmental discontinuities. This has been analytically proven for an even more general class of models than those studied in the classical period⁴. So this aspect of the results of Scheffer and van Nes¹ is not new. Rather, the novelty of their results is as follows.

Previous analytical results produced single species widely spaced along the niche axis. But Scheffer and van Nes find widely spaced clumps of species occupying very similar niches. Why the difference? Analytical work looks at the long-term equilibria of models, whereas a simulation study allows the system to be observed as it moves towards these equilibria. Scheffer and van Nes take the simulation approach, which starts out with a

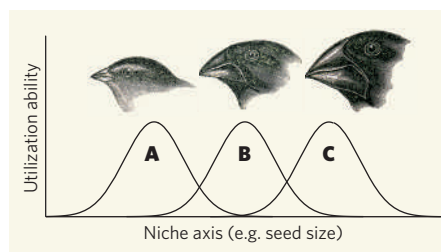


Figure 1 | How similar can species be to one another and still coexist? The question can be tackled by considering seed-eating birds (species A, B and C) in an environment that contains seeds of various sizes. For a given beak size, a bird species can optimally feed on a particular size of seed, and its feeding ability drops off for seeds that depart from this size. More generally, species are characterized by 'utilization functions', which describe the ability to exploit a particular resource as it varies along a niche axis.

large number of species along the axis and then evolves the system according to standard equations that govern competition between species. The clumps they observe are transient, and each will ultimately be thinned out to a single species. But 'ultimately' can be a very long time indeed: we now know that transient phenomena can be very long-lasting and, hence, important in ecology⁵, and such phenomena can be studied effectively only by simulation. There is also good experimental evidence for long-lasting coexistence between similar species³.

Why clumps, as opposed, for example, to a slowly thinning, uniform cloud of species along the niche axis? Consider Figure 1. In a community consisting of species A and C, where would a third species be most likely to persist successfully? Species B, positioned halfway between A and C, would be competing strongly against two species, whereas if it was in the same location as A or C it would be competing with only one. In the second case, one species would ultimately exclude the other, but — as pointed out above — this will be on a very long timescale. Of course, this argument depends on the width of the utilization distribution curves shown in Figure 1, and the distance between A and C. But this is an aspect of the self-organization of the species: they move into positions such that the void between them is an inhospitable competitive environment.

The emergence of clumps of highly similar species resonates with a proposed solution to another possible problem: the coexistence of large numbers of species in environments that do not seem to allow for much niche differentiation. Plankton and tropical forest plants are the usual examples. These organisms have a simple set of requirements: light, carbon dioxide and a few nutrients. How is it possible to carve out thousands of distinct niches from so few requirements? It has been proposed that such high numbers of species can coexist precisely because their niches are so similar that exclusion takes a very long time, perhaps on the same timescale as speciation^{6–8}.

So much for the theory: what about the data? Scheffer and van Nes present frequency histograms of species' body sizes for three data sets, which they claim show discrete clumps of species along a niche axis. For these data, body size is the niche axis — body size being the single most important variable determining a species' life history. But whereas two of their examples (aquatic beetles and phytoplankton) seem to occur in discrete clumps of species with similar body sizes, a third (American prairie birds) does not: to our eyes, these bird species look smeared out along the axis with little profound clumping. This impression illustrates the difficulty inherent in making objective judgements about clumping based simply on visual inspection of frequency histograms⁹. Statistical techniques have been developed that do, in fact, identify significant clumping among prairie birds as well as other

MATERIALS SCIENCE

Film review

Free-standing nanofilms are a wonder of membrane technology. Although it's no easy matter to produce them, once made these quasi-two-dimensional objects display fascinating behaviour, combining macroscopic surface area with nanoscopic depth.

A remarkable example is reported by Toyoki Kunitake and colleagues in *Nature Materials* (R. Vendamme *et al.* doi:10.1038/nmat1655; 2006). They have prepared an ultrathin film that is barely visible to the naked eye, but is so flexible it can be drawn through a micropipette hole 30,000 times smaller than its width (pictured).

Despite its flimsy appearance, the film can support a liquid body 70,000 times heavier than its own weight, and withstand significant deformation. It is also stable to various environmental and mechanical stresses. Even more impressively, the film breaks records for size in being several centimetres across, yet only around 35 nanometres thick.

This apparently incompatible combination of strength and thinness is a result of the film's hybrid composition. It consists of an organic polymeric network, which makes it pliable and deformable,



interpenetrated by zirconia (zirconium dioxide), which confers strength and stability. To prepare the nanofilm, the two materials are generated simultaneously from their precursors on a spin-coating plate. The chemical processes involved are quite different: the polymer forms by light-induced crosslinking of its

monomers, whereas the zirconium precursor reacts with residual traces of water in the film's polyvinyl alcohol substrate. Nevertheless, the components intertwine to give nanofilms with properties that make them useful as sensors, actuators and separation membranes.

Maria Bellantone

species⁹. This leaves the extent of overlap between statistical and ecological significance as an interesting and open question.

We can go further: on what basis did Darwin make his assertion about the discreteness of species? This question is distinct from debates about the definition of species in nature. Blackberries reproduce asexually, and it is impossible to agree on how many 'species' there are; but, nonetheless, we all know a 'blackberry' when we see one and do not wonder if it is actually a raspberry. Great tits, blue tits and coal tits are all quite distinct when considered as a set, but are surely just more-or-less continuous variants on a tit theme when compared with flamingos. Bacteria that are vastly different genetically are all called *Legionella* because they clump along the single niche axis that matters to us: they all cause Legionnaire's disease.

So what is the correct or meaningful frame of reference when thinking about the ecological nature of species? As well as providing stimulating theoretical results, Scheffer and van Nes¹ have revitalized the fundamental question of how we should look at the ecological identity of species. ■

Sean Nee and Nick Colegrave are at the Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. e-mails: sean.nee@ed.ac.uk; n.colegrave@ed.ac.uk

1. Scheffer, M. & van Nes, E. H. *Proc. Natl Acad. Sci. USA* **103**, 6230–6235 (2006).
2. Roughgarden, J. *Theory of Population Genetics and Evolutionary Ecology* (Macmillan, New York, 1979).
3. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1974).
4. Sasaki, A. J. *Theor. Biol.* **186**, 415–430 (1997).
5. Hastings, A. *Trends Ecol. Evol.* **19**, 39–45 (2004).
6. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
7. Colinvaux, P. *Ecology* **2** (Wiley, New York, 1993).
8. Nee, S. *Funct. Ecol.* **19**, 173–176 (2005).
9. Holling, C. S. *Ecol. Monogr.* **62**, 447–502 (1992).

STEM CELLS

Good, bad and reformable

Viktor Janzen and David T. Scadden

The ability of stem cells to continuously supply vast numbers of cells is magnificent, but it can be devastating if it runs amok, as in some tumours. So what makes a normal stem cell turn bad, and can it be redeemed?

The stem cell is a bit like the griffin of mythology — half lion, half eagle; grand and powerful, but potentially monstrous in effect. These essentially unspecialized cells can renew their own population while supplying cells that mature (differentiate) into the specialized cells necessary for all tissues. Although this ability to reproduce and self-renew is sublime when functioning properly, its disorder creates masses of dysfunctional replicating cells. Indeed, stem-cell-like cells have been found in a range of human tumours. Not all cancer is due to a stem cell gone bad, but some cancer-initiating cells are probably stem cells, and the rest acquire the stem-cell feature of self-renewal. This raises the troubling spectre that normal stem cells and cancer stem cells might share the molecular features essential to their nature. So attempting to treat cancer by disrupting the functions of the cancer stem cells might also disturb normal stem cells — potentially fatally.

In this issue, however, Yilmaz *et al.* (page 475)¹ and Zhang *et al.* (page 518)² report that there may be key molecular distinctions between the normal and malignant stem cell that might be of use in designing therapies that target malignant stem cells, while sparing normal stem cells.

The investigations centred on a protein called PTEN (for 'phosphatase and tensin homologue'), a known tumour suppressor and

an intracellular modulator of several major cell-signalling pathways. Notably, PTEN inhibits signalling through the AKT pathway that responds to growth factors (Fig. 1a). Growth factors bind to specific receptors on the cell surface and induce a cascade of cellular modifications in which phosphate groups are added to a series of proteins. Essentially, the activation signal is passed along the pathway like a baton in a relay race until it reaches the final 'effector' proteins that carry out the pathway response: for example, changing the expression of particular genes or halting the cell cycle. When the growth factor binds to its receptor, the enzyme PI3K is activated, and it is this step that PTEN inhibits. Activation of PI3K leads to phosphorylation and activation of the AKT protein, which in turn can potentially phosphorylate more than 9,000 proteins. Two key downstream AKT effectors, called mTOR and FOXO, are implicated in cancer development.

Yilmaz *et al.*¹ and Zhang *et al.*² used PTEN-deficient mice to examine how a lack of this protein affects cell proliferation, programmed cell death and cell localization in haematopoietic stem cells (which produce blood and immune cells) (Fig. 1b). Previous work had shown that PTEN deficiency increases the proliferation of stem or progenitor cells (a slightly more differentiated cell type) in the fetal mouse brain. It also increases self-renewal

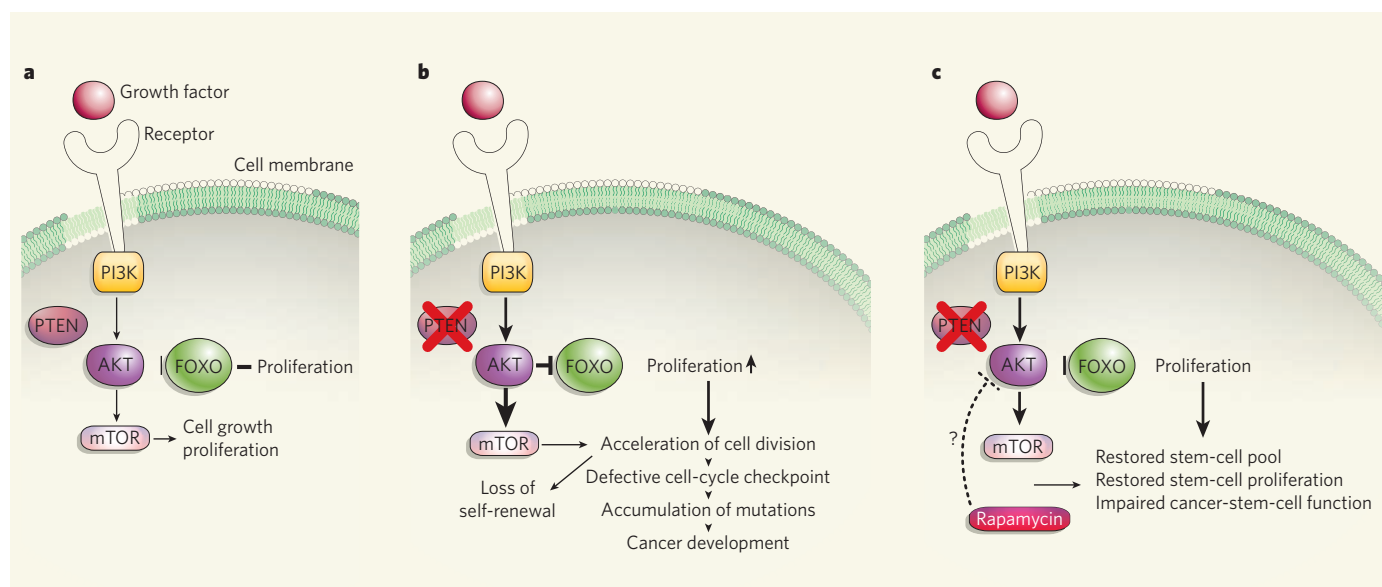


Figure 1 | PTEN regulates signal transduction and cancer development.

a, PTEN negatively regulates PI3K activity in the AKT signalling pathway that is activated by growth factors. This keeps cell division from becoming excessive. **b**, Yilmaz *et al.*¹ and Zhang *et al.*² find that, in the absence of PTEN, the proliferation of haematopoietic stem cells increases and leukaemia develops. This is presumably because hyperactivity of AKT leads to activation of mTOR with a subsequent increase in cell division. The checkpoints that would usually halt division of aberrant cells are overridden,

so that there is an accumulation of mutations and eventually malignancies develop. Hyperactive AKT would also suppress FOXO, again causing hyperproliferation. In stem cells lacking PTEN, this proliferative stress leads to exhaustion of stem-cell numbers as self-renewal is reduced. **c**, The drug rapamycin inhibits the activity of mTOR. Treatment of PTEN-deficient mice with rapamycin markedly restores stem-cell function and prevents the development of leukaemia¹. Rapamycin may also exert its effects through a direct inhibition of AKT activity.

of these cells and decreases their dependence on growth factors *in vitro*^{3,4}. In the haematopoietic system^{1,2}, the proliferation of stem cells also increased, but it was associated with three distinct effects on the homeostasis of the stem-cell pool. First, the cells changed their location, being found more commonly in the blood or spleen and becoming less able to take up residence in the bone marrow². Second, PTEN-deficient stem cells had a skewed pattern of differentiation into types of white blood cell. Third, the stem cells became progressively fewer in number^{1,2}.

If the cells are dividing more, how do they become depleted? When stem cells divide, the two daughter cells have three possible fates. Both may be stem cells, both may be differentiating cells, or one daughter may remain a stem cell while the other differentiates. The balance between the different outcomes varies depending on the circumstances. For example, during early development the stem-cell pool expands, but after organ formation is complete the stem-cell population is generally stable, with the emphasis being on producing the mature cells needed for tissue maintenance. The result of stem-cell division must therefore be modifiable, to adjust the balance between reproduction of stem cells and production of differentiating cells. So stem-cell proliferation and stem-cell expansion are not synonymous terms, and there is growing evidence that the proliferative programme is distinct from the molecular programme of self-renewal. Indeed, downstream of PTEN are molecules such as p21^{Cip1/Waf1} that influence stem-cell proliferation without necessarily

invoking self-renewal^{5,6}. The effects of PTEN on p21^{Cip1/Waf1} may be mediated by FOXO regulatory proteins that influence the expression of a range of crucial genes for stem-cell function, including cell division and tolerance of oxidative stress^{7,8}. Perturbing PTEN may thus result in non-renewing cell divisions or death of stem cells.

Yilmaz *et al.*¹ and Zhang *et al.*² found that the absence of PTEN not only caused the gradual loss of normal stem cells, but also changed their differentiation characteristics, allowing leukaemia to develop. The leukaemia included high numbers of cells that functioned as cancer stem cells¹. How can a defect that reduces self-renewal in normal stem cells also result in cancer stem cells with a capacity to self-renew?

The mechanism was not explored in these papers. But to examine whether some of the tumour-related effects are mediated through specific elements of the AKT pathway, Yilmaz *et al.*¹ looked at the role of the AKT effector mTOR. This enzyme is inhibited by the drug rapamycin (Fig. 1c), and treating the PTEN-deficient mice with rapamycin not only prevented the development of leukaemia, but also inhibited established leukaemia from progressing by profoundly reducing the numbers of cancer stem cells. These latter cells therefore remained sensitive to the pathway modulated by PTEN and targeted by rapamycin. Moreover, addition of rapamycin led to the recovery in the numbers of normal stem cells. How rapamycin, which alters only some of the many downstream effects of PTEN and AKT, could redress all the consequences of the lack of PTEN is intriguing. There are indications

that rapamycin may feed back to AKT, inhibiting its activation and thus its many downstream targets⁹. So it may be that rapamycin can alter even mTOR-independent members of the PTEN pathway, such as FOXO (Fig. 1c). If so, a drug that selectively targets a portion of the pathway might have broader beneficial effects.

Yilmaz *et al.*¹ and Zhang *et al.*² have thus shown that the loss of PTEN has distinct effects on malignant and normal stem cells that might be exploited by a therapeutic drug. This molecular defect in PTEN-deficient mice enabled the preferential emergence of diseased stem cells, but also provided a drug target to selectively enhance normal over malignant cells. The message is therefore one of hope: a single molecular defect can make a whole stem-cell pool go bad, but in this case a single targeted therapy might restore the good and impair the bad: even the griffin may be tamed. ■

Viktor Janzen and David T. Scadden are at the Center for Regenerative Medicine, Massachusetts General Hospital, Harvard Stem Cell Institute, 185 Cambridge Street, Boston, Massachusetts 02114, USA.
e-mail: scadden.david@mgh.harvard.edu

1. Yilmaz, Ö. H. *et al.* *Nature* **441**, 475–482 (2006).
2. Zhang, J. *et al.* *Nature* **441**, 518–522 (2006).
3. Groszer, M. *et al.* *Science* **294**, 2186–2189 (2001).
4. Groszer, M. *et al.* *Proc. Natl Acad. Sci. USA* **103**, 111–116 (2006).
5. Cheng, T. *et al.* *Science* **287**, 1804–1808 (2000).
6. Kippin, T. E., Martens, D. J. & van der Kooy, D. *Genes Dev.* **19**, 756–767 (2005).
7. Seoane, J. *et al.* *Cell* **117**, 211–223 (2004).
8. Essers, M. A. *et al.* *Science* **308**, 1181–1184 (2005).
9. Sarbassov, D. D. *et al.* *Mol. Cell* **22**, 159–168 (2006).

BRIEF COMMUNICATIONS

Avoidance of disease by social lobsters

These gregarious animals shun lobsters that carry a lethal virus, even when they still seem to be healthy.

Transmissible pathogens are the bane of social animals¹, so they have evolved behaviours to decrease the probability of infection^{2,3}. There is no record, however, of social animals avoiding diseased individuals of their own species in the wild. Here we show how healthy, normally gregarious Caribbean spiny lobsters (*Panulirus argus*) avoid conspecifics that are infected with a lethal virus. Early detection and avoidance of infected, though not yet infectious, individuals by healthy lobsters confers a selective advantage and highlights the importance of host behaviour in disease transmission among natural populations.

Panulirus argus virus 1 (PaV1) is a lethal pathogenic virus that infects juvenile spiny lobsters⁴. It is transmitted by physical contact and, among the smallest juveniles, through sea water. Spiny lobsters are social and share communal dens, so these modes of viral transmission could have devastating consequences in the absence of a mechanism to check its spread.

During underwater surveys of juvenile lobsters, we observed that infected lobsters rarely shared shelters with conspecifics (less than 7% shared dens and more than 93% were solitary), even though healthy lobsters generally preferred to live together (more than 56% shared dens and less than 44% were solitary).

To test whether this could be explained by healthy lobsters avoiding diseased individuals, we set up a laboratory experiment (for methods, see supplementary information) in which healthy and diseased lobsters were given a choice between an empty den and one containing either a healthy or a diseased individual. We found that the healthy lobsters avoided dens containing diseased conspecifics and preferred to share dens with other healthy lobsters. But diseased lobsters did not discriminate between dens, irrespective of whether the animal inside was sick or well (Fig. 1a). These results confirm our observations in the wild.

For an avoidance strategy to reduce the transmission of infection effectively, healthy lobsters must also start to avoid infected individuals before they actually become infectious. We therefore investigated whether there was any link between the timing of infectivity and the avoidance of diseased individuals by healthy lobsters; we found that these were remarkably coincident (Fig. 1b). Lobsters inoculated with PaV1 developed symptoms of disease after six weeks and became infectious after eight weeks (Fig. 1b). Most of the healthy

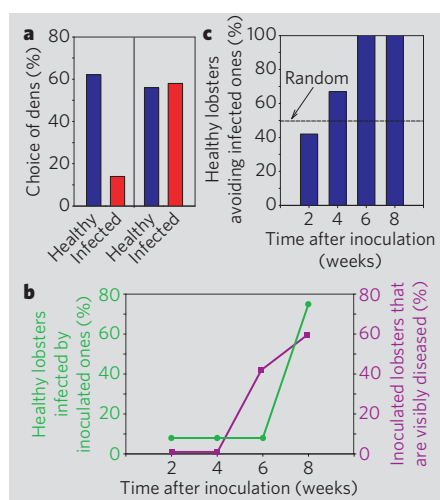


Figure 1 | Laboratory results showing avoidance by lobsters of diseased individuals and its relation to infectivity. **a**, Den selection by healthy (left panel) and infected (right panel) lobsters, showing the aversion of healthy lobsters for diseased lobsters. All lobsters were either given a choice of sheltering alone or with a healthy conspecific (blue bars), or given a choice of sheltering alone or with an infected (red bars) conspecific that was tethered in the den. **b**, Increase of infectivity over time for lobsters inoculated with PaV1 virus: green line shows the increasing percentage of healthy lobsters infected by inoculated lobsters, and purple line indicates the increasing percentage of inoculated lobsters that became visibly diseased. **c**, Cohabitation time series indicating that healthy lobsters started to avoid inoculated (but non-infectious) lobsters four weeks after inoculation and completely avoided them after six weeks, which is before they became infectious.

lobsters avoided these inoculated lobsters from four weeks after their inoculation and all of them avoided inoculated individuals at six to eight weeks after inoculation (Fig. 1c).

Other evidence (our unpublished results) indicates that avoidance of diseased conspecifics may act powerfully against the spread of disease in the wild. Laboratory confinement of healthy juvenile lobsters with lobsters infected with PaV1 results in more than 60% of them succumbing within 80 days. However, field surveys conducted since 1999 indicate that the prevalence of PaV1 in Florida's juvenile lobster population remains at less than 7% and does not correlate with lobster density (our unpublished results). The high transmissibility under laboratory conditions and the much lower prevalence in the wild, even at high density, can be reconciled if healthy, normally social lobsters thwart transmission of PaV1 by avoiding diseased lobsters.

Although infection may alter the subtle visual displays used by lobsters to establish aggregations, avoidance of diseased lobsters is likely to be chemically mediated, given that olfaction mediates dominance hierarchies, mate choice, foraging and aggregation^{5,6}. Lobsters do not have an adaptive immune system, so disease may be signified by broad-spectrum defensin compounds or by degradatory products that are associated with poor health.

Pathogens can alter host dynamics and modify communities by disrupting species interactions⁷⁻⁹, but their effect on the behaviour of uninfected hosts and on disease dynamics has been largely unexplored. Epidemiological

models have not generally included avoidance strategies^{10,11}, despite the possible fitness advantages conferred by such behaviour^{12,13}. Our findings indicate that lobsters can identify infected individuals before they become infectious and that by avoiding them they may limit disease transmission in the wild.

Donald C. Behringer*, Mark J. Butler*, Jeffrey D. Shields†

*Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529, USA
e-mail: mbutler@odu.edu

†Virginia Institute of Marine Science, Gloucester Point, Virginia 23062, USA

- Cote, I. M. & Poulin, R. *Behav. Ecol.* **6**, 159–165 (1995).
- Minchella, D. J. *Parasitology* **90**, 205–216 (1985).
- Loehle, C. *Ecology* **76**, 326–335 (1995).
- Shields, J. D. & Behringer, D. C. *Dis. Aquat. Org.* **59**, 109–118 (2004).
- Penn, D. & Potts, W. K. *Trends Ecol. Evol.* **13**, 391–396 (1998).
- Derby, C. D. *Physiol. Behav.* **69**, 203–209 (2000).
- Bushman, P. J. & Atema, J. J. *Chem. Ecol.* **26**, 883–899 (2000).
- Washburn, J. O. et al. *Science* **253**, 185–188 (1991).
- Hudson, P. & Greenman, J. *Trends Ecol. Evol.* **13**, 387–390 (1998).
- Anderson, R. M. & May, R. M. *Population Biology of Infectious Diseases* (Springer, Dusseldorf, 1982).
- McCallum, H., Barlow, N. & Hone, J. *Trends Ecol. Evol.* **16**, 295–300 (2001).
- Nevitt, G., Pentcheff, N. D., Lohman, K. J. & Zimmer, R. K. *Mar. Ecol. Prog. Ser.* **203**, 225–231 (2000).
- Rosenquist, R. & Johansson, P. *Anim. Behav.* **49**, 1039–1045 (1995).

Supplementary information accompanies this communication on Nature's website.
Received 22 February; accepted 13 April 2006.
Competing financial interests: declared none.
doi:10.1038/441421

BRIEF COMMUNICATIONS ARISING online
www.nature.com/bca see Nature contents.

CORRESPONDENCE PATTERNS

Mechanisms and models of human dynamics

Arising from: J. G. Oliveira & A.-L. Barabási *Nature* **437**, 1251 (2005)

A stochastic queue model of human behaviour developed on the basis of the distribution of timing of tasks, as studied in a sample of e-mail messages at a university¹ and in the written correspondence of Albert Einstein², may not be as simple as it seems. Although this model reproduces the apparently non-Poisson distribution of correspondence delays, its interpretation is more complicated than suggested by Barabási, who claims that humans execute their tasks based on some perceived priority, setting up queues that generate very uneven waiting-time distributions for different tasks¹. Such an explanation is intuitively appealing in the context of the queue metaphor, but does not exclude other mechanisms. By attributing delays in the correspondence to task priorities, this explanation ignores two important classes of mechanism that also contribute to the apparent distributions of task timings: the semantic content of an individual's correspondence and the social context in which this correspondence occurs.

For example, Oliveira and Barabási use Albert Einstein's two-year delay in correspondence with Theodor Kaluza as an illustration of the 'priority list' model². Although it is true that Einstein did not communicate Kaluza's work to the Prussian Academy of Sciences until two years after Kaluza first sent his manuscript to Einstein, this was not due to a perceived unimportance of the task by Einstein. Rather, on 21 April 1919, Einstein wrote a prompt reply acknowledging the receipt of Kaluza's manuscript just a few days earlier³. On 29 May 1919, Einstein suggested that Kaluza should revise his current manuscript to resolve a flaw in his unified field theory, and encouraged Kaluza to submit the work for publication in *Mathematische Zeitschrift*, even offering "to put in a good word" with the editors³. To help Kaluza, he enclosed a manuscript of his own that related the electromagnetic and gravitational fields⁴.

Kaluza attempted to address the flaw pointed out by Einstein, but ultimately was unable to, referring to it as a "serious difficulty" in the final version published in 1921 (ref. 5). Thus, the delay between Einstein's receipt of Kaluza's original manuscript and its final publication was probably due to the difficulty that Einstein's question posed for Kaluza, and not to the low priority that Einstein assigned either to replying to Kaluza or in recommending his manuscript for publication.

Even when questions posed in letters are easy to answer, responses and their consequent actions may be delayed because they are part of a many-person conversation and require information from other interlocutors. For

example, the distribution observed for e-mail delays¹ may be due to groups of individuals engaging in shared conversations, in which individuals send messages not according to personal priorities, but rather to their collective attributes such as kinship, hierarchy and affiliation. Earlier analysis of the e-mail data set used by Barabási revealed significant many-person conversations, as reflected in the temporal coherence of three-person e-mail exchanges⁶. These social structures and their temporal correspondence properties were due to groups of individuals who were affiliated with academic departments, comprised non-academic committees, and shared common foreign nationalities⁶. Their existence is consistent with the idea that the mechanism underlying the apparent non-Poisson distribution of e-mail timings is related to the collective properties of the inherently social groups that produce them.

In all communications, both the semantic content of tasks, as in the correspondence between Einstein and Kaluza, as well as their social context, as in the correspondence of e-mail users, are important factors in the mechanics of human interaction. The stochastic queue provides a useful model and an elegant metaphor for the interdependence of

human tasks, but this does not mean that it is determined by individual priority lists. Indeed, the priority queue metaphor could be replaced by a semantic ladder or social web, depending on the nature of the variable (personal priority, task difficulty or social relevance, respectively), that parameterizes the model's response time. Determination of the mechanism(s) underlying the apparent distribution of human task timings requires precise investigation of the relationships among the semantic content of tasks, task priorities as perceived by individuals, and the collective attributes of the social groups that govern human behaviour.

Alex Kentsis

Department of Physiology & Biophysics, Mount Sinai School of Medicine, New York University, New York, New York 10029, USA
e-mail: alex.kentsis@mssm.edu

1. Barabási, A. L. *Nature* **435**, 207–211 (2005).
2. Oliveira, J. G. & Barabási, A. L. *Nature* **437**, 1251 (2005).
3. Einstein, A. *The Collected Papers of Albert Einstein* (Princeton Univ. Press, Princeton, New Jersey, 1987).
4. Einstein, A. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* 349–356 (1919).
5. Kaluza, T. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* **54**, 966–972 (1921).
6. Eckmann, J. P., Moses, E. & Sergi, D. *Proc. Natl Acad. Sci. USA* **101**, 14333–14337 (2004).

doi:10.1038/nature04901

CORRESPONDENCE PATTERNS

Barabási & Oliveira reply

Replying to: A. Kentsis *Nature* **441**, doi:10.1038/nature04901 (2006)

Kentsis notes that the response time to an e-mail or a letter depends on the semantic content of the correspondence, as well as the social context in which the communication arises¹. We would add that it also depends on deadlines, the time dependence of priorities and the dropping of past-deadline messages², making human response dynamics sufficiently complicated that no simple model could fully account for it^{3–6}. However, the advantage of the proposed modelling framework is that most of these effects can be incorporated into it, and their impact on the queuing process can be systematically evaluated. Addressing some of these additional mechanisms, including those suggested by Kentsis, requires information that is beyond reach for most researchers at this point.

Thanks to legitimate concerns about privacy, we do not have access to the text of e-mail messages, which limits our ability to analyse

both the semantic and the social context in an automatic fashion. But it is not impossible to envisage advances in these directions, given that important e-mail services (such as gmail) analyse message content for the purpose of placing advertisements, as well as using algorithms developed in the context of interruptibility research to evaluate both content and social context⁷. The proposed models thus represent the starting point for a quantitative understanding of human dynamics.

The comment raises an important question: given that so little information is available for several key parameters, why are the proposed models so successful at capturing the observed response-time statistics, for both e-mails³ and letters⁴? To understand this, we note that message content, social context and deadlines are important only for assigning the correct priority to each message. However, it can be shown that the response-time distribution is inde-

pendent of the priority distribution³. Therefore, to predict the response statistics correctly, we do not need to know the correct priority of each message, but any priority distribution will lead to the same response-time distribution.

Given this fact, we can distinguish two goals of potential interest. One is predicting the response statistics of the overall communication pattern, at which queuing models are excellent, even without information on social context, message content and the precise value of the priorities assigned to each message. The second is predicting the precise response time to a specific message, which will require an evaluation of all factors that have an impact on the specific message. In other words, without knowing all the circumstances of the

Kaluza–Einstein correspondence, including Einstein's and Kaluza's competing responsibilities, it would be impossible to foresee the two-year delay in the publication of Kaluza's famous paper. Yet we can accurately predict the fraction of letters that waited for one day, one month or one year on Einstein's desk before he would write a response, as — amazingly — the response-time distribution is completely independent of the individual priorities assigned to a letter. This universality and independence of details is the real strength of the proposed modelling framework.

J. G. Oliveira*†, **A.-L. Barabási***‡

*Department of Physics and Center for Complex Networks Research, University of Notre Dame, Indiana 46556, USA

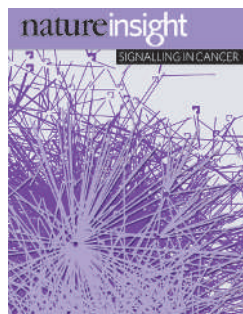
e-mail: alb@nd.edu

†Departamento de Física, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

‡Center for Cancer Systems Biology, Dana Farber Cancer Institute, Harvard University, Boston, Massachusetts 02115, USA

1. Kentsis, A. *Nature* **441**, doi:10.1038/nature04901 (2006).
2. Vázquez, A. et al. *Phys. Rev. E* **73**, 036127 (2006).
3. Barabási, A.-L. *Nature* **435**, 207–211 (2005).
4. Oliveira, J. G. & Barabási, A.-L. *Nature* **437**, 1251 (2005).
5. Vázquez, A. *Phys. Rev. Lett.* **95**, 248701 (2005).
6. Hidalgo, C. *Physica A* (in the press); preprint at <http://arxiv.org/abs/cond-mat/0512278> (2005).
7. Horvitz, E., Kadie, C. M., Paek, T. & Hovel, D. *Commun. ACM* **46**, 52–59 (2003).

doi:10.1038/nature04902

**Cover illustration**

Adapted from a network graph of protein-protein interactions in

J.-F. Rual *et al.* *Nature* **437**, 1173–1178 (2005).

Editor, *Nature*

Philip Campbell

Insights Publisher

Sarah Greaves

Insights Editor

Lesley Anson

Production Editors

Nell Boase

Rebecca Craven

Senior Art Editor

Martin Harrison

Art Editor

Nik Spencer

Layouts

Jane Wright

Sponsorship

Claudia Banks

Claire Hines

Production

Jocelyn Hilton

Marketing

Robin Brown

Editorial Assistant

Laura Shaw

SIGNALLING IN CANCER

The past 30 years have led from the discovery of the first cancer-causing gene, or 'oncogene', to the emergence of a new generation of cancer therapies — those targeted at specific signalling molecules.

The signalling pathways controlling cell growth and differentiation are almost invariably altered in cancer. These interconnected pathways are being deciphered, but understanding the alterations that lead to cancer and correcting them is a substantial challenge. The reviews in this Insight discuss the molecular circuitry regulating several key cellular processes, and illustrate how defining the signalling mechanisms is aiding the development of therapies.

Among the key pathways are those controlling cell proliferation, which coordinate a response to the cellular environment, with the mTOR kinase as a critical node. Tumour development is influenced by infections and inflammation, and the complex role of the nuclear factor- κ B transcription factors is being unravelled. Expansion of tumour cells depends on nutrient supply and vascularization, which is orchestrated by the transcription factor known as HIF. And the metastatic spread of primary tumours to other organs is facilitated by many signalling pathways; exploring their functional contributions has just begun.

Evaluating signalling molecules as drug targets is important for prioritizing research, even though we cannot predict the success of drugs in the clinic. Still, with several inhibitors of signalling molecules now approved for clinical use, and more in the pipeline, there is reason to celebrate 30 years of oncogene research. We hope these reviews provide a glimpse of recent excitements. Thanks are due to the authors for their contributions and to reviewers for their input.

We are pleased to acknowledge the financial support of Genentech and Pfizer in producing this Insight. As always, *Nature* carries sole responsibility for editorial content and peer review.

Alex Eccleston, Senior Editor

Ritu Dhand, Chief Biology Editor

REVIEWS

424 **Ras, PI(3)K and mTOR signalling controls tumour cell growth**

R. J. Shaw & L. C. Cantley

431 **Nuclear factor- κ B in cancer development and progression**

M. Karin

437 **Hypoxia signalling in cancer and approaches to enforce tumour regression**

J. Pouyssegur, F. Dayan & N. M. Mazure

444 **New signals from the invasive front**

G. Christofori

451 **Validating cancer drug targets**

J. D. Benson, Y.-N. P. Chen, S. A. Cornell-Kennon, M. Dorsch, S. Kim, M. Leszczyniecka, W. R. Sellers & C. Lengauer

457 **Mechanisms of drug inhibition of signalling molecules**

J. S. Sebolt-Leopold & J. M. English

nature
insight

Ras, PI(3)K and mTOR signalling controls tumour cell growth

Reuben J. Shaw¹ & Lewis C. Cantley²

All eukaryotic cells coordinate cell growth with the availability of nutrients in their environment. The mTOR protein kinase has emerged as a critical growth-control node, receiving stimulatory signals from Ras and phosphatidylinositol-3-OH kinase (PI(3)K) downstream from growth factors, as well as nutrient inputs in the form of amino-acid, glucose and oxygen availability. Notably, components of the Ras and PI(3)K signalling pathways are mutated in most human cancers. The preponderance of mutations in these interconnected pathways suggests that the loss of growth-control checkpoints and promotion of cell survival in nutrient-limited conditions may be an obligate event in tumorigenesis.

Three decades of basic cancer research have revealed that mutations in components of signalling pathways that control cell growth in primitive metazoans underlie tumour initiation in mammals. The Ras, PI(3)K and mTOR (mammalian target of rapamycin) signalling pathways form an intersecting biochemical network that, when mutated, drives cell growth in a manner unrestricted by environmental cues. Ultimately, these pathways drive tumorigenesis through the coordinated phosphorylation of proteins that directly regulate protein synthesis, cell-cycle progression and metabolism, and of transcription factors that regulate the expression of genes involved in these processes^{1,2}.

The basic elements of this biochemical network are outlined in Fig. 1. Growth factors activate receptor tyrosine kinases (RTKs), which then activate two key signal-transduction components: the small GTPase Ras and the lipid kinase PI(3)K. Most human tumours harbour activating mutations in these master regulators (K-ras, H-ras, N-ras, the p110 α PI(3)K subunit and RTKs), or inactivating mutations in negative regulators of these proteins (phosphatase and tensin homologue (PTEN) and neurofibromin 1 (NF1)). More recent genomic sequencing efforts have also revealed oncogenic mutations in several downstream components of these pathways³.

The serine/threonine kinase mTOR is a highly conserved integrator of both mitogenic and nutrient inputs in yeast and mammalian cells, and has been shown to control cell growth in response to various environmental cues. Recent discoveries indicate that the Ras and PI(3)K pathways converge to activate mTOR to stimulate cell growth. Notably, several tumour suppressors of previously unknown function (tuberous sclerosis complex 1 (TSC1, also known as hamartin), TSC2 (tuberin) and serine/threonine protein kinase 11 (LKB1)) have recently been shown to attenuate mTOR signalling under nutrient-poor conditions. Accordingly, inactivation of *TSC1*, *TSC2* or *LKB1*, or of the aforementioned PTEN and NF1, results in familial cancer syndromes with shared clinical features (phakomatosis). So, this ancestral network, which evolved to ensure that cell proliferation occurs only under environmentally favourable conditions, has been exploited by cancer cells to promote growth and survival under inappropriate conditions^{2,4}.

Recent pharmaceutical efforts in developing kinase inhibitors have resulted in a number of agents designed to inhibit the kinase components of these signalling pathways (mTOR, PI(3)K, RTKs, Raf and AKT).

Although these pathways are among the most thoroughly studied in molecular cancer research, recent advances indicate that the components have a much more complex role in cellular and organismal physiology than was previously appreciated. This new insight has led to additional therapeutic targets and has provided a framework that should facilitate the design of combined therapies aimed at tumours with specific genetic lesions.

Here we highlight recent findings that more fully illuminate how these pathways contribute to tumorigenesis and discuss some of the issues faced in designing rational therapies for them.

Ras and Raf

Roughly one decade after the identification of *RAS* as an important human oncogene, the first critical direct effector of Ras in mammalian cells was identified: the Raf-1 serine/threonine kinase¹. This discovery linked Ras to the ERK (extracellular-signal-regulated kinase) mitogen-activated protein kinase (MAPK) pathway and stimulated interest in inhibitors of this pathway for chemotherapeutic intervention in Ras-dependent tumours. More recently, activating mutations in one of the Raf isoforms, B-raf, have been found in a large (>60%) fraction of human malignant melanomas, and in some colon, thyroid and lung tumours⁵. Tumour-associated mutations in B-raf have been found in some cases to cause constitutive heterodimerization with the related family member c-raf (also called Raf-1), which then activates downstream ERK signalling⁶. Of particular interest here is recent evidence linking this pathway to activation of the mTOR protein kinase (see below). However, it is not clear whether the Raf-ERK pathway is the most critical mediator of Ras-dependent tumorigenesis in all cell types; several other Ras effectors have been identified in the past ten years⁷ (Box 1).

Prolonged activation of Ras can occur in tumours by mechanisms that do not involve mutations in Ras. One of the earliest-identified tumour-suppressor genes was *NF1*, which encodes a GAP (GTPase-activating protein) for Ras⁸. Loss of NF1 results in the accumulation of Ras in the GTP-bound state due to a decreased rate of GTP hydrolysis. Interestingly, none of the other 12 genes in the human genome that encode RasGAPs has been shown to be inactivated in human cancer. Why NF1 seems to be unique among RasGAPs in suppressing human neoplasia remains unknown. However, an RNA-interference screen for

¹Dulbecco Laboratory for Cancer Research, The Salk Institute, 10010 N. Torrey Pines Road, La Jolla, California 9203, USA. ²Department of Systems Biology, Harvard Medical School, and Division of Signal Transduction, Beth Israel Deaconess Medical Center, 300 Brookline Avenue, Boston, Massachusetts 02215, USA.

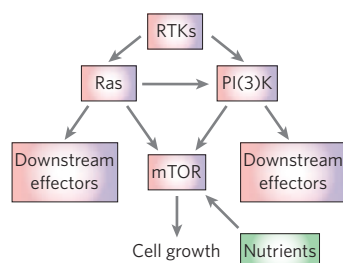


Figure 1 | Ancient growth-control pathways. Mitogens acting through receptor tyrosine kinases (RTKs) activate Ras GTPase and phosphatidylinositol-3-OH kinase (PI(3)K). Each of these proteins then activates a number of downstream effectors. One effector critical to cell growth that is stimulated by both Ras and PI(3)K is the mTOR (mammalian target of rapamycin) kinase. In addition to these growth-factor inputs, mTOR activity is controlled by the availability of nutrients (glucose, amino acids and oxygen).

novel suppressors of Ras activity identified a transcription factor that promoted expression of the RasGAP RASAL1 (ref. 9). In addition to loss of RasGAP function, Ras activity is reported to be elevated through reduced expression of a conserved microRNA (let-7) that targets the 3' untranslated region of H-ras, N-ras and K-ras messenger RNAs¹⁰.

PI(3)K pathway

The PI(3)K pathway has been implicated in cancer since its discovery as an enzymatic activity associated with viral oncoproteins 20 years ago, but in the past 5 years it has become apparent that it is one of the most frequently targeted pathways in all sporadic human tumours, with estimates suggesting that mutation in one or another PI(3)K pathway component accounts for up to 30% of all human cancers². PI(3)K is activated by both RTKs and Ras, and in turn activates several downstream signalling pathways through the generation of the lipid second messenger phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P₃; Box 2). In particular, the AKT family (also known as protein kinase B or PKB) of serine/threonine kinases has emerged as a critical target of PI(3)K in human cancer.

Some of the first direct evidence for PI(3)K deregulation in human cancer was the discovery of amplification of the genes encoding the p110 α PI(3)K catalytic subunit and AKT2 in ovarian, breast and pancreatic cancer². Mutations in the gene encoding the PI(3)K regulatory subunit p85 α were found in some primary colon and ovarian tumours³. Strikingly, a large-scale effort to sequence exons of PI(3)K genes from human tumours revealed clustered regions of point mutations in the p110 α catalytic subunit in 20–30% of breast, colon, brain and gastric tumours examined¹¹. Investigations of the most frequent p110 α tumour mutations have shown that they enhance PI(3)K activity and drive cell transformation¹². Further sequencing of exons from genes encoding components of the PI(3)K pathway in colon tumours revealed point mutations in AKT2 and 3-phosphoinositide-dependent protein kinase 1 (PDK1 or PDPK1), and amplifications of AKT2 and insulin receptor substrate 2 (IRS2), in addition to mutations in PIK3CA (p110 α) and PTEN¹³. Furthermore, the *TCL1* oncogene from human T-cell leukaemias with 14q32-1 translocations has been shown to activate AKT¹⁴, suggesting that there may be other, undiscovered cancer genes that regulate this pathway.

Despite the surprisingly high rates of activating mutations in p110 α , loss of the PTEN lipid phosphatase still appears to be the most common mechanism of activation of the PI(3)K pathway in human cancers. PTEN catalyses removal of the D3 phosphate from PtdIns(3,4,5)P₃ (the reverse of the reaction catalysed by PI(3)K) to limit and ultimately terminate PI(3)K signalling in cells¹⁵. Germline mutations in PTEN have been found in a collection of dominantly inherited cancer syndromes with overlapping symptoms, including Cowden's disease and Bannayan-Zonana syndrome. Sporadic mutations of PTEN are found in a high percentage of many tumour types, including breast, ovarian and colon cancers and glioblastoma¹⁶. PTEN is now thought to be the second most commonly mutated tumour suppressor in humans, after p53.

The PI(3)K–AKT pathway is also negatively regulated by protein serine/threonine phosphatases. PHLPP, a newly identified phosphatase containing a lipid-binding PH (pleckstrin homology) domain, specifically dephosphorylates Ser 473 of AKT¹⁷. Two related PHLPP genes are located in chromosomal regions mutated in colon and breast cancer, respectively. Inhibition of expression of PHLPP has been shown to drive AKT activation in mammalian and *Drosophila* cells and to promote tumour growth of xenografts in nude mice. Future studies of this interesting protein should reveal its mechanism of regulation and its overall relevance to human cancer.

AKT couples cell growth and cell survival to metabolism

Through the phosphorylation of a diverse set of substrates, AKT regulates four intersecting biological processes: cell survival, cell-cycle progression, cell growth and cell metabolism. The AKT substrates that mediate some of these biological processes have been identified, but it is unlikely that we know all the critical AKT substrates. Here we describe a few substrates with the best-known relevance to human cancer.

Glycogen synthase kinase-3 (GSK-3) is a highly conserved protein kinase that is inhibited by AKT phosphorylation¹⁸. GSK-3 controls a number of critical cell-cycle events through the phosphorylation of cell-cycle regulators such as c-Myc, cyclin D1 and cyclin E. GSK-3 also phosphorylates transcription factors that govern cell fate and differentiation, including c-Jun, β -catenin, GLI, Notch, Snail and sterol-regulatory-element-binding transcription factor 1 (SREBP1). In general, the phosphorylation of proteins by GSK-3 results in impaired function, in some cases due to turnover of the phosphorylated substrate by the creation of a phospho-docking site for ubiquitin ligases such as FBW7 or FWD1 (ref. 19). So, by inactivating GSK-3, AKT may enhance the functions of some of these transcription factors.

AKT controls cell-cycle progression through several further substrates. It has been implicated in a radiation-damage checkpoint^{20,21}, in which it may operate in part through direct phosphorylation of the Chk1 protein kinase²² or through phosphorylation of MDM2, a human oncogene product that degrades the p53 tumour suppressor²³. AKT also controls cell survival through its inactivation of the pro-apoptotic protein BAD^{24,25} and its activation of the I κ B kinase (IKK)–NF κ B (nuclear factor- κ B) pathway (see the review in this issue by Karin, page 431).

The FOXO family of forkhead transcription factors is a set of highly conserved substrates of AKT. AKT phosphorylation of FOXO proteins creates a 14-3-3 binding site, resulting in their inactivation through sequestration in the cytoplasm. Conversely, FOXO proteins are activated under stress conditions, accumulating in the nucleus and driving the expression of pro-apoptotic genes as well as stress-response genes, such as superoxide dismutase (SOD2), that contribute to lifespan extension in lower metazoans²⁶. Although FOXOs are well established as being critical in the regulation of lifespan and metabolism downstream of AKT, the role for FOXO in tumour suppression has not yet been clearly defined. Importantly, FOXOs are disrupted by translocations in some rhabdomyosarcomas and leukaemias²⁷.

For more than a decade, PI(3)K and AKT have been implicated in the activation of the mTOR protein kinase. One critical target of AKT that regulates mTOR is the TSC2 tumour suppressor protein, tuberlin²⁸. TSC2 (which is discussed in more detail below) had been previously identified as a hamartoma syndrome gene in humans and as a gene affecting cell size in flies. Tuberlin negatively regulates mTOR signalling through its ability to act as a GAP for the Ras-like GTPase Rheb, and AKT activation was found to circumvent this inhibition⁴ (Box 3).

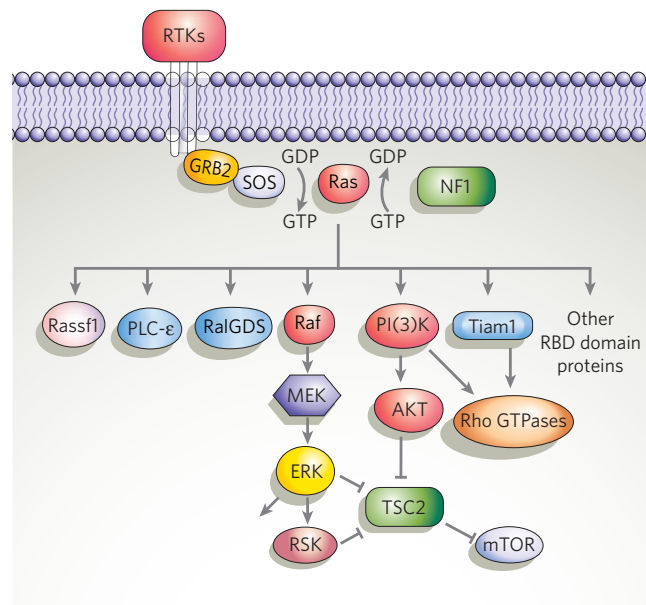
Mutational activation of the PI(3)K–AKT pathway also has a critical role in maintaining cell metabolism in conditions of limiting growth factors. AKT has been shown to stimulate cell growth and ATP production by regulating the activity and expression of key glycolytic enzymes as well as the cellular uptake of glucose and other nutrients. It has been proposed that AKT directly stimulates glycolysis through the phosphorylation of 6-phosphofructo-2-kinase (PFK2), and fatty-acid synthesis through the phosphorylation of ATP citrate lyase²⁹. Of particular interest, recent studies have shown that AKT and mTOR can mediate

Box 1 | Ras signalling

Activating oncogenic mutations in three Ras isoforms and several downstream effectors are found in human cancer (red in the figure); inactivating mutations in tumor suppressors in this pathway shown in green. Several other Ras effectors are required for Ras-induced tumorigenesis in genetic mouse models (blue).

Receptor tyrosine kinase (RTK) activation is known to signal to Ras through a number of evolutionarily conserved steps. On binding ligands, RTKs dimerize, activating their kinase activity and leading to phosphorylation of tyrosine sites in their cytoplasmic tails and on adaptor proteins such as those of the IRS (insulin receptor substrate) family. The phosphorylated tyrosine residues serve as docking sites for various adaptor proteins containing SH2 or PTB domains such as GRB2. These adaptor proteins recruit further effectors, including Ras guanine-nucleotide-exchange factors (GEFs) such as SOS, that stimulate the exchange of GDP for GTP on Ras. On binding GTP, Ras undergoes a conformational change that allows downstream effectors to bind to its 'switch I' region. GTPase-activating proteins (GAPs) such as p120^{GAP} and neurofibromin 1 (NF1) also bind to activated Ras and catalyse GTP hydrolysis, thereby returning Ras to the inactive GDP-bound state. Of the roughly 21 members of the Ras subfamily of small GTPases, only H-ras, K-ras and N-ras have been shown to undergo frequent mutations in human cancer. The tumour mutations found in Ras lock it into the GTP-bound state such that it remains constitutively active.

All known Ras effectors share a region of weak homology, called the Ras-binding domain (RBD), that interacts with the effector loop of Ras. Effectors of Ras can be divided into three subgroups based on similarities in RBDs and on response to effector loop mutations of Ras⁷. Raf family protein kinases (c-raf, A-raf, and B-raf) make up one class. In a complex with scaffolding proteins such as CNK (connector enhancer of kinase suppressor of Ras 1) and KSR (kinase suppressor of Ras 1), activated Raf phosphorylates dual-specificity MAPK/ERK (MEK) kinases, which then phosphorylate and activate the ERK1 (extracellular-signal-regulated kinase 1) and ERK2 mitogen-activated protein kinases (MAPKs). ERK phosphorylates a wide variety of substrates that mediate cell growth and cell-cycle entry. In addition to transcription factors (for example Myc, Elk1 and c-Fos), ERK phosphorylates and activates the RSK (ribosomal protein S6 kinase) and MNK (MAPK-interacting serine/threonine kinase) families



of kinases. The substrates of these kinases that contribute to cell-growth control have recently been extensively reviewed⁶⁹.

The phosphatidylinositol-3-OH kinase (PI(3)K) family comprises a second major class of Ras effectors. The third class of Ras effectors is a diverse set of proteins whose RBD domains are also called RA (Ras-associating) domains. This class includes GEFs that stimulate the Ral family of GTPases to activate the p38 MAPK pathway (for example, RalGDS) and a novel phospholipase C family member (PLC-ε)⁷⁰. Finally, the RBD-containing Rho GTPase family GEF Tiam1 (T-cell lymphoma invasion and metastasis 1) serves as one route to activate the Rac and Rho GTPases downstream of Ras⁷¹. Although it is largely unknown whether effectors other than Raf or PI(3)K have a critical role in human cancer, mice lacking RalGDS, PLC-ε or Tiam1 showed reduced tumour incidence in Ras-dependent models of tumorigenesis⁷²⁻⁷⁴.

activation of the HIF-1α (hypoxia-inducible factor-1α) transcription factor, which increases expression of the glucose transporter GLUT1 and glycolytic enzymes, ultimately leading to increased glucose uptake and glycolysis (see the review in this issue by Pouyssegur *et al.*, page 437).

Two distinct mTOR complexes

In yeast and in mammals, there are two distinct and mutually exclusive TOR complexes, each composed of TOR, a common regulatory subunit called LST8, and at least a third subunit that specifies the downstream substrates. In mammals, the substrate-defining subunits are raptor (the mTORC1 complex) and rictor (mTORC2). Whereas mTORC1 complexes are strongly inhibited by rapamycin, mTORC2 is not affected by the drug. In yeast, the TORC1 complex couples transcription, ribosome biogenesis, translation initiation, nutrient uptake and autophagy to the availability of nutrients, whereas TORC2 controls cell polarity and the spatial control of cell growth⁴. Most of the work done in mammalian cells until a year ago focused on the rapamycin-sensitive mTORC1 complex. Two well-characterized substrates of the mTORC1 complex that control translation and cell growth are the 4EBP1 family of proteins and the S6 protein kinases (S6K1 and S6K2) (Box 3).

Adding an extra layer of complexity is the recent finding that the rapamycin-insensitive mTORC2 complex is required for phosphorylation of the hydrophobic motif at Ser 473 of AKT³⁰. The AKT Ser 473 site is analogous to the hydrophobic motif site in S6K that is known to be regulated by the mTORC1 complex. Loss of TORC2 also results in loss of phosphorylation of the hydrophobic motif of another related kinase, protein kinase Cα, in both yeast and mammalian cells³¹. Genetically, RNA interference (RNAi) for TORC2 components leads to a complete loss of phosphorylation at the AKT Ser 473 site in *Drosophila*, in

Dictyostelium and in a variety of mammalian cells^{30,32,33}. Finally, purified mTORC2, but not mTORC1, from growth-factor-stimulated cells resulted in *in vitro* phosphorylation of the Ser 473 site of AKT³¹. These results suggest that the catalytic activity of the TOR kinase directly phosphorylates hydrophobic motifs in AGC family kinases, although the possibility that the TORC1 and TORC2 complexes act as scaffolds to drive auto-*cis* or auto-*trans* phosphorylation of AGC family members or to suppress the dephosphorylation of these kinases has not been excluded.

These results suggest a complex interplay between the two different TORC complexes during growth-factor-mediated cell growth in mammalian cells. On growth factor stimulation, PI(3)K activation results in recruitment of AKT to the plasma membrane, in which phosphorylation at AKT Ser 473 is mediated by the mTORC2 complex and phosphorylation at Thr 308 is mediated by PDK1. The mTORC2-activated AKT then phosphorylates and inactivates tuberlin, resulting in increased mTORC1 activity. An open question is whether rictor or raptor is limiting for complex formation. Because only the raptor-mTOR complex is inhibited by rapamycin, it remains to be seen whether inhibitors specific to rictor-mTOR or inhibitors of the catalytic site of mTOR will be useful in cancer treatment.

Activation of mTOR by the ERK pathway

As mentioned above, growth factors signal to mTORC1 complexes not only through AKT but also through the ERK pathway. Tuberlin is a direct substrate of ERK (Ser 664)^{34,35} and is also a substrate of the downstream ribosomal protein S6 kinase (RSK; Ser 1798)³⁶. ERK-dependent phosphorylation has been reported to negatively regulate TSC2 function by blocking its interaction with TSC1 (ref. 35), and RSK-dependent

phosphorylation has been reported to inhibit the ability to turn off Rheb³⁶. Interestingly, in NF1-deficient cells and tumours, endogenous levels of activated Ras were shown to induce mTOR activation through PI(3)K and AKT³⁷. In the context of tumorigenesis, it is possible that the relative contributions of each of these pathways to mTOR activation will depend on the cell type and/or growth factor.

Inhibition of mTORC1 by the LKB1-AMPK pathway

In addition to growth-factor-mediated stimulation, mTORC1 activity depends on the availability of nutrients such as glucose, oxygen and amino acids. Recently, a number of proteins that regulate mTOR in response to nutrient availability have been discovered, and here we focus on those with known connections to cancer (for more extensive coverage, see Wullschlegel *et al.*⁴). LKB1 is a tumour suppressor that is mutated in the familial cancer disorder Peutz–Jeghers syndrome, as well as in a large percentage of sporadic lung adenocarcinomas^{38,39}. LKB1 encodes a threonine kinase that was shown genetically and biochemically to be the direct activating kinase for the AMP-activated protein kinase (AMPK), a mediator of cellular and organismal metabolism^{40–42}. When intracellular ATP levels drop and AMP levels rise, such as under conditions of hypoxia or glucose deprivation, AMP directly binds a subunit of AMPK, causing a conformational shift that exposes the activation-loop threonine, which is then phosphorylated by LKB1 (ref. 43). In non-transformed cultured cells, conditions that elevate intracellular AMP cause a complete inhibition of mTORC1 activity without affecting PI(3)K–AKT signalling. However, in cells that lack LKB1, mTORC1 remains active because the AMPK checkpoint is defective^{44,45}. Consistent with these cell-culture studies, hamartomas derived from *Lkb1*-heterozygous mice exhibit enhanced mTORC1 signalling, in contrast to adjacent normal epithelium⁴⁵. Furthermore, acute genetic deletion of LKB1 in the liver of adult mice in the absence of overt stress leads to direct activation of mTORC1 signalling⁴⁶. These findings suggest that the central role of AMPK in the inhibition of mTOR under normal physiological conditions has been underestimated because tissue-culture cells are grown in conditions of supraphysiological levels of glucose, oxygen and growth factors.

AMPK inhibits mTOR at least in part by the direct phosphorylation of tuberlin (Box 3), and cells that lack tuberlin retain activated mTORC1 under conditions of low glucose as well as hypoxia^{47,48}. Strikingly, the clinical symptoms of Peutz–Jeghers syndrome (germline mutations in *LKB1*) overlap with those of patients with Cowden's disease (germline mutations in *PTEN*) and with tuberous sclerosis (germline mutations in *TSC1* or *TSC2*). All three diseases are marked by the development of histologically similar hamartomas, although the tissues of origin vary. A common biochemical link between these three diseases is that the initiating genetic lesion results in the activation of mTORC1. The finding that mTOR signalling is enhanced in four different human hamartoma syndromes indicates that rapamycin analogues may be useful therapeutically for these diseases.

Pathway circuitry dictates biological outcome

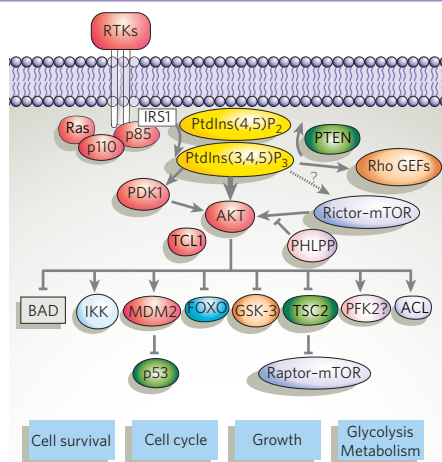
In addition to the acute shut-off mechanisms for growth-factor signalling that were discussed above (for example RasGAPs, PtdIns(3,4,5)P₃ phosphatases and AKT phosphatases), prolonged stimulation of growth-factor signalling pathways can result in additional time-delayed shut-off mechanisms. These time-delayed negative-feedback loops provide further protection from uncontrolled cell growth, but they can also introduce complications in predicting appropriate targets for pharmaceutical intervention in cancer. Negative feedback signals for Ras have been reported by means of ERK-mediated phosphorylation⁴⁹ and transcriptional events⁵⁰, although the full circuitry and contexts in which this feedback may have a role in cancer remains to be elucidated.

Similarly, prolonged activation of the PI(3)K pathway results in multiple forms of negative-feedback regulation⁵¹. Several years ago, prolonged insulin signalling or the presence of excessive nutrients, particularly in the context of animal models of obesity and diabetes, were shown to downmodulate PI(3)K and AKT activation. Subsequently, in both flies and mammals, TSC-deficient cells were found to have extremely low levels of AKT activation, even after growth-factor stimulation, despite high mTORC1 activity^{52–54}. Interestingly, these effects were inhibited by prolonged treatment with rapamycin. This suppression of PI(3)K signalling seemed to be mediated at more than one level. First, S6K

Box 2 | PI(3)K and AKT signalling

Three major classes of PI(3)Ks exist in mammals, but so far only the class IA PI(3)K subgroup has been implicated in cancers. The class IA PI(3)Ks are heterodimers of p85 family regulatory subunits and p110 family catalytic subunits. They are capable of phosphorylating the D3 position of the inositol ring of the plasma membrane lipid phosphatidylinositol-4,5-bisphosphate to generate the second messenger phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)P₃). The p85 subunits each contain two SH2 domains that bind to specific phosphotyrosine residues in the tails of activated RTKs and/or associated adaptors such as those of the IRS family. So, the PI(3)K heterodimer is recruited to the membrane to generate local pools of PtdIns(3,4,5)P₃. In addition to being associated with and activated by RTKs, the p110 catalytic subunit can also bind directly to activated Ras, and this seems to contribute to the activation of PI(3)K. PtdIns(3,4,5)P₃ acts as a membrane-bound second messenger by localizing a subset of signalling proteins with PH (pleckstrin homology) domains to the membrane, where they become activated and initiate downstream signalling events.

Of all the PH-domain-containing proteins activated by PI(3)K, the best characterized



and most directly implicated in human cancer are members of the AKT family of protein serine/threonine kinases (AKT1, AKT2 and AKT3). AKT is activated by two distinct phosphorylation events, both of which depend on PI(3)K. 3-Phosphoinositide-dependent protein kinase 1 (PDK1 or PDPK1) phosphorylates the activation-loop Thr 308 of AKT after the binding of PtdIns(3,4,5)P₃ to the PH domains of both kinases. Elegant studies with *Dpdk1* knock-in mice have shown that the

PH domain of PDK1 is required for its ability to activate AKT, but that it is dispensable for the PDK1-dependent activation of other AGC family protein kinases⁷⁵. Conversely, mutations in the 'PIF pocket' domain of PDK1 do not affect AKT activation but abolish phosphorylation of the other PDK1 substrates. Therefore, of the many kinases that PDK1 phosphorylates, only AKT phosphorylation by PDK1 depends on PI(3)K. The identity of the kinase, 'PDK2', responsible for phosphorylation of the second critical site in AKT, Ser 473, in the hydrophobic motif has long been sought and remains controversial. Many different candidate kinases have been reported to mediate this event *in vitro*, with few convincing genetic or physiologically relevant data *in vivo*. This year, it was reported that a complex of the mTOR (mammalian target of rapamycin) kinase bound to a unique adaptor called rictor mediates the phosphorylation of AKT Ser 473 *in vivo*³⁰.

AKT controls cell survival, cell cycle, cell growth and metabolism through phosphorylation of a number of key substrates (only some of which are shown in the figure). Proteins whose genes are mutationally activated in human cancer are shown in red; those inactivated are in green (as in Box 1).

Box 3 | mTORC1 signalling

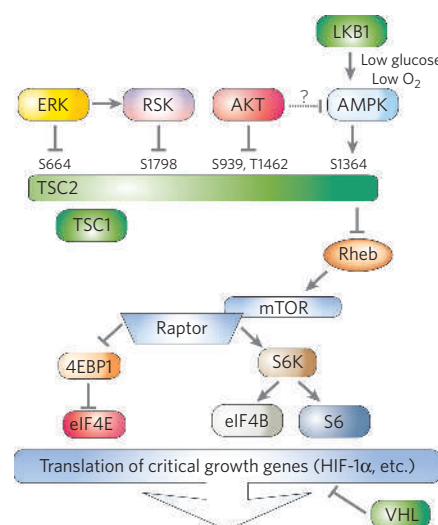
The mTOR kinase is an integrator of growth-factor and nutrient signals. Growth-factor signalling through Ras-ERK and PI(3)K-AKT activates mTORC1, whereas low nutrient availability (for example, low glucose or hypoxia) inhibits mTORC1, in part through the LKB1-AMPK (AMP-activated protein kinase) pathway. As in Boxes 1 and 2, proteins whose genes are mutationally activated in human cancer are shown in red in the figure; those inactivated are in green.

Upstream components of the raptor-mTORC1 pathway were initially discovered through classical cancer genetics. The TSC2 tumour suppressor, tuberlin, and its obligate binding partner, hamartin (TSC1), are mutated in a familial tumour syndrome called tuberous sclerosis complex (TSC). TSC patients are predisposed to widespread benign tumours termed hamartomas in kidney, lung, brain and skin. The carboxy terminus of TSC2 contains a conserved GAP domain that acts on the small Ras-like GTPase Rheb. So, analogously to NF1 and Ras, loss of TSC2 results in Rheb activation, leading to tumour development. Rheb associates with mTOR complexes^{76,77}, but how GTP binding to Rheb activates the raptor-mTOR complex is poorly understood.

One way in which the mTORC1 complex is activated by growth factors and inhibited by nutrient shortage seems to be through the phosphorylation of tuberlin (TSC2) by several upstream kinases. Signals that inhibit TSC2, and thus activate mTORC1, include the kinases ERK, RSK and AKT, all of which directly phosphorylate TSC2 *in vivo*. AKT directly phosphorylates TSC2 on a number of sites, several of which are conserved

between mammals and *Drosophila*, although the requirement of these sites for AKT-mediated regulation of mTOR remains an area of active investigation. Conversely, AMPK phosphorylation of TSC2 activates its ability to inhibit mTORC1, even in the presence of active ERK and AKT, but the underlying mechanism is unknown^{45,47}. Furthermore, each of these kinases may have additional substrates in the mTORC1 pathway, and the relative importance of each of the conserved TSC2 phosphorylation sites is being investigated at present. Finally, AKT has been reported to crosstalk and inhibit AMPK, further stimulating mTOR activation^{78,79}. Phosphorylation sites are numbered from human TSC2 (accession number P49815).

Downstream of the mTORC1 complex are its two well-characterized substrates: 4EBP1 and the p70 ribosomal S6 kinase (S6K). Phosphorylation of 4EBP1 by mTORC1 suppresses its ability to bind and inhibit the translation-initiation factor eIF4E. mTORC1 mediates phosphorylation of S6K at a threonine residue in a hydrophobic motif at the C terminus of the kinase domain, and this phosphorylation allows the recruitment and subsequent phosphorylation and activation of S6K by PDK1 (the same kinase that activates AKT; see Box 2). A specific motif (the TOS motif) found in both 4EBP1 and S6K mediates direct binding of these proteins to raptor, allowing them to be phosphorylated in the mTORC1 complex. Recently, mechanistic details of how mTOR regulates the assembly of translation-initiation complexes through a number of ordered phosphorylation events were discovered⁸⁰. How mTOR controls



translation by means of these substrates has been reviewed^{81,82}.

mTORC1-dependent translation is known to control a number of specific cell-growth regulators, including the HIF-1α (hypoxia-inducible factor-1α) transcription factor, which in turn drive diverse processes including cell growth, glycolysis and angiogenesis, all contributing to enhanced tumorigenesis. Interestingly, HIF-1α protein degradation is independently negatively regulated by the von Hippel-Lindau (VHL) tumour suppressor⁸³, providing another link between this pathway and cancer, and another genetic setting in tumours in which mTOR inhibitors may prove efficacious⁸⁴.

and perhaps mTOR itself were found to phosphorylate the IRS1 and IRS2 adaptor proteins directly on multiple sites, leading to acute inhibition of their abilities to activate PI(3)K. Second, the levels of IRS1 and IRS2 proteins and mRNAs were found to be reduced in response to prolonged mTOR activation, and this effect could be reversed by prolonged rapamycin treatment or by the introduction of RNAi for S6K1 and S6K2 (refs 32, 54).

This downregulation of PI(3)K-AKT signalling by an mTOR-S6K-dependent process has recently been shown to be important in both metabolic disease and cancers. When wild-type mice are placed on a high-fat diet for months, mTOR-S6K activity is enhanced in insulin-sensitive tissues, whereas PI(3)K-AKT activity in response to insulin is suppressed. Importantly, mice deficient in S6K1 were refractory to this obesity-induced insulin resistance, supporting a model in which persistent mTOR-S6K activity causes insulin resistance⁵⁵.

Hyperactivation of mTOR-S6K signalling and inhibition of AKT signalling was also observed in benign haemangiomas that occur in *Tsc2*^{+/-} mice⁵⁶. Strikingly, *Tsc2*^{+/-} mice that were also heterozygous for *Pten* showed increased AKT signalling in haemangiomas, and this correlated with a transition from benign to more aggressive and invasive pathology. Indeed, the nearly complete inhibition of PI(3)K-AKT activity in TSC-deficient tumours might explain why these tumours rarely become metastatic. Although increased mTOR activity fuels the initial cellular overgrowth, the subsequent inhibition of AKT signalling to other downstream targets (GSK-3, FOXO, MDM2, and so on) restricts tumour progression. This may explain the observation that sporadic human tumours commonly contain mutations in PTEN or PI(3)K but rarely have mutations in TSC1 or TSC2. Acquired secondary mutations in TSC1 or TSC2

might actually dampen tumour cell growth through the feedback inhibition of PI(3)K-AKT. Finally, although evidence for the downregulation of PI(3)K signalling through S6K-mediated IRS1/IRS2 inhibition is compelling, other mechanisms are probably involved, including inhibition of RTK expression⁵⁷ and perhaps competition between the rictor-mediated and raptor-mediated assembly of mTOR complexes⁵⁸.

Implications for therapeutic response

These findings have strong implications for cancer therapeutic strategies. Rapamycin-based mTOR inhibitors have been introduced into several clinical trials in the past couple of years⁵⁹. The existence of the S6K-based negative-feedback loop means that prolonged rapamycin treatment will probably lead to enhanced PI(3)K-AKT activation in some tumours, as has been observed in several tumour cell lines grown in culture, and in tissues from patients in clinical trials of rapamycin analogues^{60,61}. Depending on the other mutations found in the tumour, this hyperactivation of PI(3)K-AKT signalling could make the tumour more aggressive. These findings suggest that rapamycin analogues should be used in combination with inhibitors of PI(3)K-AKT signalling or inhibitors of upstream RTKs that are driving PI(3)K activation.

Upregulation of the critical targets downstream of mTORC1 in cancerous cells provides another mechanism for circumventing the inhibition of tumour growth by rapamycin. Lowe and colleagues showed that a combination of rapamycin and wortmannin inhibited lymphoid tumours originating from *MYC* and *AKT* transgenes, supporting the ideas expressed above. However, they found that overexpression of eIF4E (the binding partner of the mTORC1-regulated protein 4EBP1) avoided the rapamycin inhibition of growth⁶². These data indicate that in

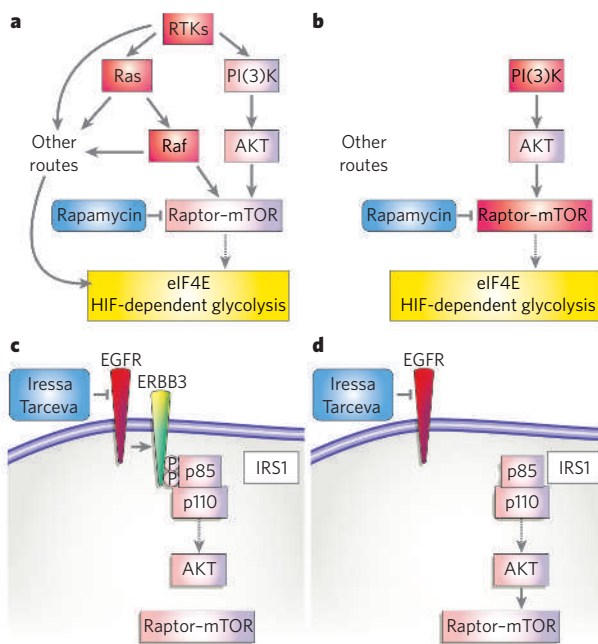


Figure 2 | Pathway circuitry dictates therapeutic response. **a**, For tumours with defined genetic lesions, the ability to overcome a given targeted therapeutic lies in whether or not they need to acquire a secondary genetic mutation to overcome the effect of the drug on critical downstream biochemical effectors that are required for continued tumour cell growth, or whether they can simply upregulate existing alternative routes that lead to effectors already expressed in those cells. So, the drug places selection pressure to ramp up existing bypass routes. If there are no such routes to the critical downstream effectors, a specific mutation to upregulate those alternative routes or bypass the drug are required. In this example, a critical target for tumour cell growth and survival is the activation of eIF4E and HIF. Tumours with initiating mutations in RTKs, Ras or Raf have multiple routes to signal to eIF4E and HIF, so blocking mTOR with rapamycin does not inhibit these tumours. **b**, In contrast, tumours with initiating lesions in PI(3)K or more direct regulators of mTOR (such as LKB1 and TSC) do not have alternative routes to activate eIF4E and HIF. Hence these tumours show greater response to rapamycin. **c**, Similarly, the expression and use of specific adaptor proteins that enhance certain arms of pathway signalling will dictate the therapeutic response. In the example shown, human lung tumours expressing epidermal-growth-factor receptor (EGFR) are targeted with anti-EGFR drugs such as Iressa or Tarceva. In tumours expressing the ERBB3 heterodimerization partner, EGFR efficiently enhances PI(3)K activation through a number of PI(3)K-binding sites in ERBB3. **d**, In tumours that lack ERBB3, PI(3)K is still activated by a number of other routes, including adaptors such as insulin receptor substrate 1 (IRS1).

this tumour type, the critical function of the mTORC1 complex required for cell growth is activation of eIF4E. Because eIF4E transcription is controlled by multiple signalling inputs, resistance to rapamycin might develop by a variety of paths that will depend on the genetic background of the tumour (Fig. 2a, b).

More broadly, identifying the primary mechanisms by which the PI(3)K pathway is activated in a given tumour should facilitate the choice of therapeutic intervention. For example, it was recently found that expression of ERBB3 predicted the therapeutic response of human non-small-cell lung tumour cell lines to the epidermal-growth-factor receptor (EGFR) inhibitor Iressa (gefitinib)⁶³. Human ERBB3 contains multiple p85-binding sites, allowing potent PI(3)K recruitment and activation in response to EGFR activation and heterodimerization with ERBB3. Lung tumours that express ERBB3 were more sensitive to Iressa, as the drug robustly inhibited PI(3)K activation in those cells. In cells lacking ERBB3, PI(3)K was activated by other routes (for example, IRS1) that were not affected by EGFR inhibitors, and so these cells were able to keep PI(3)K active and were insensitive to growth inhibition by the drug (Fig. 2c, d).

For tumours bearing Ras mutations, defining the critical downstream effector in each given tissue context is another important future goal. In some settings, such as melanoma, Raf has been assumed to be the critical Ras effector, as Raf mutations are common and mutually exclusive with N-ras mutations⁵. In contrast, mutational data suggest that activating mutations in Ras family members and loss of PTEN are mutually exclusive in some tumour settings in both mice and humans (for example, skin and endometrial tumours)⁶⁴, so in these tissues PI(3)K may be the critical Ras effector in tumorigenesis. Conversely, mutations in Raf and PTEN are found together in human melanoma^{65,66}, and in a recent examination of the PI(3)K pathway in colon cancer, 27 of the 36 tumours showing activating point mutations in the p110α PI(3)K subunit also had activating Ras mutations¹².

There is also clear evidence for both synergy and redundancy between Raf and PI(3)K-mediated signalling on specific biochemical effectors. Activation of mTORC1 activity can be mediated by distinct phosphorylation events on tuberin by ERK and RSK, as well as by AKT. The cell-death effector BAD is also inhibited by both of these pathways⁶⁷. In addition, given the close similarity of the RSK and AKT kinase domains, it is possible that both can phosphorylate the same sites on some target proteins, such as GSK-3 (ref. 68) or even tuberin³⁵. These redundancies may be exploited by tumour cells faced with targeted therapeutics. Again, combination therapies will probably be the key to reducing the initial fraction of tumour cells that can bypass a given selection pressure through upregulation of an alternative signalling route.

Perspectives and future directions

Understanding the full circuitry of a signalling pathway, including feedback loops, is a requirement for an understanding of the biological consequences of perturbing that pathway. This is true in therapeutic cancer intervention as inhibitors of mTOR or other PI(3)K signalling components enter clinical trials. Defining the components of the signalling pathway that initiate tumour formation is critical for developing an intervention strategy. Tumours appear to become 'addicted' to these initiating events, perhaps because negative-feedback loops have repressed all other avenues of growth. Thus, drugs that target the underlying perturbation are more likely to be effective and less likely to have side effects than drugs that act as general growth suppressors. The hope is that clinical trials with agents that target specific components of signal-transduction networks will be designed to tease out the subset of patients whose tumours are known to be driven by the component of interest. Defining the combinations of inhibitors that are likely to be effective on tumours with specific combinations of genetic lesions will be an even greater challenge. However, on the basis of our increasing understanding of these networks, there is great hope that this strategy will ultimately be highly effective. As more drugs that target specific components of signal-transduction pathways become available and as we increase our knowledge of the complexity of these signalling networks, the burden of selecting the correct drug combinations for each individual cancer patient will ultimately shift to the pathologist, who must identify the underlying defect in each tumour. This will require new diagnostic technologies and will be a major challenge over the next decade.

1. McCormick, F. Signalling networks that cause cancer. *Trends Cell. Biol.* **9**, M53–M56 (1999).
2. Luo, J., Manning, B. D. & Cantley, L. C. Targeting the PI3K–Akt pathway in human cancer: rationale and promise. *Cancer Cell* **4**, 257–262 (2003).
3. Bader, A. G., Kang, S., Zhao, L. & Vogt, P. K. Oncogenic PI3K deregulates transcription and translation. *Nature Rev. Cancer* **5**, 921–929 (2005).
4. Wulfschleger, S., Loewith, R. & Hall, M. N. TOR signaling in growth and metabolism. *Cell* **124**, 471–484 (2006).
5. Garnett, M. J. & Marais, R. Guilty as charged: B-Raf is a human oncogene. *Cancer Cell* **6**, 313–319 (2004).
6. Garnett, M. J., Rana, S., Paterson, H., Barford, D. & Marais, R. Wild-type and mutant B-Raf activate C-Raf through distinct mechanisms involving heterodimerization. *Mol. Cell* **20**, 963–969 (2005).
7. Rodriguez-Viciana, P., Sabatier, C. & McCormick, F. Signaling specificity by Ras family GTPases is determined by the full spectrum of effectors they regulate. *Mol. Cell. Biol.* **24**, 4943–4954 (2004).
8. Cichowski, K. & Jacks, T. NF1 tumor suppressor gene function: narrowing the GAP. *Cell* **104**, 593–604 (2001).

9. Kolfschoten, I. G. *et al.* A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell* **121**, 849–858 (2005).
10. Johnson, S. M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635–647 (2005).
11. Samuels, Y. *et al.* High frequency of mutations of the *PIK3CA* gene in human cancers. *Science* **304**, 554 (2004).
12. Samuels, Y. *et al.* Mutant *PIK3CA* promotes cell growth and invasion of human cancer cells. *Cancer Cell* **7**, 561–573 (2005).
13. Parsons, D. W. *et al.* Colorectal cancer: mutations in a signalling pathway. *Nature* **436**, 792 (2005).
14. Teitell, M. A. The TCl1 family of oncoproteins: co-activators of transformation. *Nature Rev. Cancer* **5**, 640–648 (2005).
15. Maehama, T. & Dixon, J. E. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* **273**, 13375–13378 (1998).
16. Cantley, L. C. & Neel, B. G. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. *Proc. Natl Acad. Sci. USA* **96**, 4240–4245 (1999).
17. Gao, T., Furnari, F. & Newton, A. C. PHLPP: a phosphatase that directly dephosphorylates Akt, promotes apoptosis, and suppresses tumor growth. *Mol. Cell* **18**, 13–24 (2005).
18. Frame, S. & Cohen, P. GSK3 takes centre stage more than 20 years after its discovery. *Biochem. J.* **359**, 1–16 (2001).
19. Minella, A. C. & Clurman, B. E. Mechanisms of tumor suppression by the SCF^{Fbw7}. *Cell Cycle* **4**, 1356–1359 (2005).
20. Kandel, E. S. *et al.* Activation of Akt/protein kinase B overcomes a G₂/M cell cycle checkpoint induced by DNA damage. *Mol. Cell. Biol.* **22**, 7831–7841 (2002).
21. Lee, C., Kim, J. S. & Waldman, T. PTEN gene targeting reveals a radiation-induced size checkpoint in human cancer cells. *Cancer Res.* **64**, 6906–6914 (2004).
22. Puc, J. *et al.* Lack of PTEN sequesters CHK1 and initiates genetic instability. *Cancer Cell* **7**, 193–204 (2005).
23. Mayo, L. D. & Donner, D. B. The PTEN, Mdm2, p53 tumor suppressor-oncoprotein network. *Trends Biochem. Sci.* **27**, 462–467 (2002).
24. Datta, S. R. *et al.* Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery. *Cell* **91**, 231–241 (1997).
25. del Peso, L., Gonzalez-Garcia, M., Page, C., Herrera, R. & Nunez, G. Interleukin-3-induced phosphorylation of BAD through the protein kinase Akt. *Science* **278**, 687–689 (1997).
26. Accilli, D. & Arden, K. C. FoxOs at the crossroads of cellular metabolism, differentiation, and transformation. *Cell* **117**, 421–426 (2004).
27. So, C. W. & Cleary, M. L. Common mechanism for oncogenic activation of MLL by forkhead family proteins. *Blood* **101**, 633–639 (2003).
28. Manning, B. D. & Cantley, L. C. United at last: the tuberous sclerosis complex gene products connect the phosphoinositide 3-kinase/Akt pathway to mammalian target of rapamycin (mTOR) signalling. *Biochem. Soc. Trans.* **31**, 573–578 (2003).
29. Hammerman, P. S., Fox, C. J. & Thompson, C. B. Beginnings of a signal-transduction pathway for bioenergetic control of cell survival. *Trends Biochem. Sci.* **29**, 586–592 (2004).
30. Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor–mTOR complex. *Science* **307**, 1098–1101 (2005).
31. Sarbassov, D. D. *et al.* Rictor, a novel binding partner of mTOR, defines a rapamycin-insensitive and raptor-independent pathway that regulates the cytoskeleton. *Curr. Biol.* **14**, 1296–1302 (2004).
32. Hresko, R. C. & Mueckler, M. mTOR/RICTOR is the Ser473 kinase for Akt/PKB in 3T3-L1 adipocytes. *J. Biol. Chem.* **280**, 40406–40416 (2005).
33. Lee, S. *et al.* TOR complex 2 integrates cell movement during chemotaxis and signal relay in Dictyostelium. *Mol. Biol. Cell* **16**, 4572–4583 (2005).
34. Ballif, B. A. *et al.* Quantitative phosphorylation profiling of the ERK/p90 ribosomal S6 kinase-signaling cassette and its targets, the tuberous sclerosis tumor suppressors. *Proc. Natl Acad. Sci. USA* **102**, 667–672 (2005).
35. Ma, L., Chen, Z., Erdjument-Bromage, H., Tempst, P. & Pandolfi, P. P. Phosphorylation and functional inactivation of TSC2 by Erk implications for tuberous sclerosis and cancer pathogenesis. *Cell* **121**, 179–193 (2005).
36. Roux, P. P., Ballif, B. A., Anjum, R., Gygi, S. P. & Blenis, J. Tumor-promoting phorbol esters and activated Ras inactivate the tuberous sclerosis tumor suppressor complex via p90 ribosomal S6 kinase. *Proc. Natl Acad. Sci. USA* **101**, 13489–13494 (2004).
37. Johannessen, C. M. *et al.* The NF1 tumor suppressor critically regulates TSC2 and mTOR. *Proc. Natl Acad. Sci. USA* **102**, 8573–8578 (2005).
38. Hemminki, A. *et al.* A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature* **391**, 184–187 (1998).
39. Sanchez-Cespedes, M. *et al.* Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. *Cancer Res.* **62**, 3659–3662 (2002).
40. Hawley, S. A. *et al.* Complexes between the LKB1 tumor suppressor, STRADA/β and MO25α/β are upstream kinases in the AMP-activated protein kinase cascade. *J. Biol.* **2**, 28 (2003).
41. Woods, A. *et al.* LKB1 is the upstream kinase in the AMP-activated protein kinase cascade. *Curr. Biol.* **13**, 2004–2008 (2003).
42. Shaw, R. J. *et al.* The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress. *Proc. Natl Acad. Sci. USA* **101**, 3329–3335 (2004).
43. Hardie, D. G. New roles for the LKB1-AMPK pathway. *Curr. Opin. Cell Biol.* **17**, 167–173 (2005).
44. Corradetti, M. N., Inoki, K., Bardeesy, N., DePinho, R. A. & Guan, K. L. Regulation of the TSC pathway by LKB1: evidence of a molecular link between tuberous sclerosis complex and Peutz-Jeghers syndrome. *Genes Dev.* **18**, 1533–1538 (2004).
45. Shaw, R. J. *et al.* The LKB1 tumor suppressor negatively regulates mTOR signaling. *Cancer Cell* **6**, 91–99 (2004).
46. Shaw, R. J. *et al.* Deletion of the LKB1 kinase in liver impairs AMPK activation, glucose homeostasis, and therapeutic effects of metformin. *Science* **310**, 1642–1646 (2005).
47. Inoki, K., Zhu, T. & Guan, K. L. TSC2 mediates cellular energy response to control cell growth and survival. *Cell* **115**, 577–590 (2003).
48. Liu, L. *et al.* Hypoxia-induced energy stress regulates mRNA translation and cell growth. *Mol. Cell* **21**, 521–531 (2006).
49. Dougherty, M. K. *et al.* Regulation of Raf-1 by direct feedback phosphorylation. *Mol. Cell* **17**, 215–224 (2005).
50. Macrae, M. *et al.* A conditional feedback loop regulates Ras activity through EphA2. *Cancer Cell* **8**, 111–118 (2005).
51. Manning, B. D. Balancing Akt with S6K: implications for both metabolic diseases and tumorigenesis. *J. Cell Biol.* **167**, 399–403 (2004).
52. Harrington, L. S. *et al.* The TSC1-2 tumor suppressor controls insulin-PI3K signaling via regulation of IRS proteins. *J. Cell Biol.* **166**, 213–223 (2004).
53. Radimerski, T., Montagne, J., Hemmings-Mieszczak, M. & Thomas, G. Lethality of *Drosophila* lacking TSC tumor suppressor function rescued by reducing dS6K signaling. *Genes Dev.* **16**, 2627–2632 (2002).
54. Shah, O. J., Wang, Z. & Hunter, T. Inappropriate activation of the TSC/Rheb/mTOR/S6K cassette induces IRS1/2 depletion, insulin resistance, and cell survival deficiencies. *Curr. Biol.* **14**, 1650–1656 (2004).
55. Um, S. H. *et al.* Absence of S6K1 protects against age- and diet-induced obesity while enhancing insulin sensitivity. *Nature* **431**, 200–205 (2004).
56. Manning, B. D. *et al.* Feedback inhibition of Akt signaling limits the growth of tumors lacking Tsc2. *Genes Dev.* **19**, 1773–1778 (2005).
57. Zhang, H. *et al.* Loss of Tsc1/Tsc2 activates mTOR and disrupts PI3K-Akt signaling through downregulation of PDGFR. *J. Clin. Invest.* **112**, 1223–1233 (2003).
58. Hay, N. The Akt–mTOR tango and its relevance to cancer. *Cancer Cell* **8**, 179–183 (2005).
59. Granville, C. A., Memmott, R. M., Gills, J. J. & Dennis, P. A. Handicapping the race to develop inhibitors of the phosphoinositide 3-kinase/Akt/mammalian target of rapamycin pathway. *Clin. Cancer Res.* **12**, 679–689 (2006).
60. Sun, S. Y. *et al.* Activation of Akt and eIF4E survival pathways by rapamycin-mediated mammalian target of rapamycin inhibition. *Cancer Res.* **65**, 7052–7058 (2005).
61. O'Reilly, K. E. *et al.* mTOR inhibition induces upstream receptor tyrosine kinase signaling and activates Akt. *Cancer Res.* **66**, 1500–1508 (2006).
62. Wendel, H. G. *et al.* Survival signalling by Akt and eIF4E in oncogenesis and cancer therapy. *Nature* **428**, 332–337 (2004).
63. Engelman, J. A. *et al.* ErbB-3 mediates phosphoinositide 3-kinase activity in gefitinib-sensitive non-small cell lung cancer cell lines. *Proc. Natl Acad. Sci. USA* **102**, 3788–3793 (2005).
64. To, M. D., Perez-Losada, J., Mao, J. H. & Balmain, A. Crosstalk between Pten and Ras signaling pathways in tumor development. *Cell Cycle* **4**, 1185–1188 (2005).
65. Daniotti, M. *et al.* BRAF alterations are associated with complex mutational profiles in malignant melanoma. *Oncogene* **23**, 5968–5977 (2004).
66. Tsao, H., Goel, V., Wu, H., Yang, G. & Haluska, F. G. Genetic interaction between NRAS and BRAF mutations and PTEN/MMAC1 inactivation in melanoma. *J. Invest. Dermatol.* **122**, 337–341 (2004).
67. She, Q. B. *et al.* The BAD protein integrates survival signaling by EGFR/MAPK and PI3K/Akt kinase pathways in PTEN-deficient tumor cells. *Cancer Cell* **8**, 287–297 (2005).
68. Ding, Q. *et al.* Erk associates with and primes GSK-3β for its inactivation resulting in upregulation of β-catenin. *Mol. Cell* **19**, 159–170 (2005).
69. Roux, P. P. & Blenis, J. ERK and p38 MAPK-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol. Mol. Biol. Rev.* **68**, 320–344 (2004).
70. Cullen, P. J. Ras effectors: buying shares in Ras plc. *Curr. Biol.* **11**, R342–R344 (2001).
71. Lambert, J. M. *et al.* Tiam1 mediates Ras activation of Rac by a PI(3)K-independent mechanism. *Nature Cell Biol.* **4**, 621–625 (2002).
72. Gonzalez-Garcia, A. *et al.* RalGDS is required for tumor formation in a model of skin carcinogenesis. *Cancer Cell* **7**, 219–226 (2005).
73. Bai, Y. *et al.* Crucial role of phospholipase Cε in chemical carcinogen-induced skin tumor development. *Cancer Res.* **64**, 8808–8810 (2004).
74. Malliri, A. *et al.* Mice deficient in the Rac activator Tiam1 are resistant to Ras-induced skin tumours. *Nature* **417**, 867–871 (2002).
75. McManus, E. J. *et al.* The *in vivo* role of PtdIns(3,4,5)P₃ binding to PDK1 PH domain defined by knockin mutation. *EMBO J.* **23**, 2071–2082 (2004).
76. Long, X., Lin, Y., Ortiz-Vega, S., Yonezawa, K. & Avruch, J. Rheb binds and regulates the mTOR kinase. *Curr. Biol.* **15**, 702–713 (2005).
77. Smith, E. M., Finn, S. G., Tee, A. R., Browne, G. J. & Proud, C. G. The tuberous sclerosis protein TSC2 is not required for the regulation of the mammalian target of rapamycin by amino acids and certain cellular stresses. *J. Biol. Chem.* **280**, 18717–18727 (2005).
78. Hahn-Windgassen, A. *et al.* Akt activates the mammalian target of rapamycin by regulating cellular ATP level and AMPK activity. *J. Biol. Chem.* **280**, 32081–32089 (2005).
79. Horman, S. *et al.* Insulin antagonizes ischemia-induced Thr172 phosphorylation of AMP-activated protein kinase α-subunits in heart via hierarchical phosphorylation of Ser485/491. *J. Biol. Chem.* **281**, 5335–5340 (2006).
80. Holz, M. K., Ballif, B. A., Gygi, S. P. & Blenis, J. mTOR and S6K1 mediate assembly of the translation preinitiation complex through dynamic protein interchange and ordered phosphorylation events. *Cell* **123**, 569–580 (2005).
81. Fingar, D. C. & Blenis, J. Target of rapamycin (TOR): an integrator of nutrient and growth factor signals and coordinator of cell growth and cell cycle progression. *Oncogene* **23**, 3151–3171 (2004).
82. Hay, N. & Sonenberg, N. Upstream and downstream of mTOR. *Genes Dev.* **18**, 1926–1945 (2004).
83. Kim, W. Y. & Kaelin, W. G. Role of VHL gene mutation in human cancer. *J. Clin. Oncol.* **22**, 4991–5004 (2004).
84. Thomas, G. V. *et al.* Hypoxia-inducible factor determines sensitivity to inhibitors of mTOR in kidney cancer. *Nature Med.* **12**, 122–127 (2006).

Acknowledgements We thank K. Cichowski, K. Lamia and J. Blenis for critical reading of the manuscript, and we apologize to many colleagues whose work could only be cited indirectly because of space limitations.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to R.J.S. (shaw@salk.edu) or L.C.C. (lcantley@hms.harvard.edu).

Nuclear factor- κ B in cancer development and progression

Michael Karin¹

Nuclear factor- κ B (NF- κ B) transcription factors and the signalling pathways that activate them are central coordinators of innate and adaptive immune responses. More recently, it has become clear that NF- κ B signalling also has a critical role in cancer development and progression. NF- κ B provides a mechanistic link between inflammation and cancer, and is a major factor controlling the ability of both pre-neoplastic and malignant cells to resist apoptosis-based tumour-surveillance mechanisms. NF- κ B might also regulate tumour angiogenesis and invasiveness, and the signalling pathways that mediate its activation provide attractive targets for new chemopreventive and chemotherapeutic approaches.

A link between inflammation and cancer was suspected well before the discovery of NF- κ B or other transcription factors. Although Virchow suggested in the nineteenth century that chronic inflammation might give rise to malignancy¹, the link between inflammation and cancer was not widely understood until more recent times^{1,2}. This 're-discovery' can be attributed, in part, to epidemiological studies that identified chronic infections and inflammation as major risk factors for various types of cancer. Collectively, underlying infections and inflammation are linked to 15–20% of all cancer deaths³. For instance, chronic infections with hepatitis B virus (HBV) and hepatitis C virus (HCV) are major risk factors for hepatocellular carcinoma (HCC), whereas infections with *Helicobacter pylori* are associated with most gastric cancers⁴. Chronic inflammatory bowel diseases (IBDs), such as ulcerative colitis (UC), are thought to increase the risk of colorectal cancer by approximately 1% per year⁵, and chronic airway irritation and inflammation caused by airborne particles and tobacco smoke are likely to be important promoters of lung carcinogenesis⁶. Although epidemiological studies are an excellent source of new working hypotheses, they only underline correlations and do not establish causal relationships or mechanistic links. Thus, the evidence listed above and elsewhere^{2,7,8} begs the question of how chronic inflammation influences tumour development and progression.

In trying to provide an answer to this question, one needs to consider the cellular processes that contribute to the emergence of neoplasia and its malignant progression. These processes were elegantly summarized by Hanahan and Weinberg as self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of apoptosis, limitless replicative potential, tissue invasion and metastasis, and sustained angiogenesis⁹. As recently discussed, inflammation and NF- κ B can affect most of these processes^{7,10} (Fig. 1). To minimize the reiteration of recent reviews on inflammation and cancer, this article focuses on the role of a single transcription factor, NF- κ B, in linking these pathophysiological processes. Although a role for NF- κ B in cancer development and progression has been suspected^{11,12}, this hypothesis has rested mainly on circumstantial evidence, such as the occurrence of constitutively active NF- κ B in many types of cancer. Only recently has solid genetic and biochemical evidence for a causative role of NF- κ B in malignant conversion and progression been obtained. As discussed below, depending on the cell type in which it acts, NF- κ B can either promote or inhibit

carcinogenesis. Such information, available from various mouse models of cancer in which NF- κ B activation has been blocked by genetic means, is crucial for the clinical development of NF- κ B inhibitors as cancer therapeutics.

How infection and inflammation affect cancer development

Cancer is a chronic disease that is caused by defective genome-surveillance and signal-transduction mechanisms⁹. If infection and inflammation enhance tumour development, they must do so through signal-transduction mechanisms that influence factors involved in either malignant conversion or cancer surveillance. In general, unless they carry their own oncogenes, toxins or growth factors, infectious organisms affect the host through pattern-recognition receptors (PRRs), most commonly those that belong to the Toll-like receptor (TLR) family^{13,14}. PRR engagement activates numerous signal-transduction pathways that target several transcription factors, which control the expression of genes encoding cytokines, chemokines and enzymes that regulate innate and adaptive immune responses^{14,15}. In turn, some of these polypeptides activate receptors that further propagate and amplify the inflammatory response, which is a particular manifestation of a much broader innate immune response (Fig. 2). Although engagement of TLRs and receptors for proinflammatory cytokines, such as tumour-necrosis factor- α (TNF- α) and interleukin-1 (IL-1), leads to activation of many important signalling pathways, it is well accepted that the central role in inflammation and innate immunity is played by the NF- κ B group of transcription factors¹⁶. The details of NF- κ B regulation are outlined in Box 1.

It is noteworthy that many of the receptors and cytokines mentioned above have also been linked to carcinogenic processes. Genetic polymorphisms in a *TLR* gene cluster are associated with high risk for prostate cancer¹⁷, whereas polymorphisms in the *IL1B* promoter that enhance IL-1 β production are associated with increased risk of gastric cancer¹⁸. Another proinflammatory cytokine involved in cancer is TNF- α . Despite being named for its ability to induce tumour necrosis, which is an activity that is mostly mediated through increased vascular permeability and subsequent vascular collapse, there is ample evidence that TNF- α acts as a tumour promoter in several models of experimental cancer^{19–23}.

Of all the different signalling pathways activated by inflammation and infection, NF- κ B might be the most important component of the

¹Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, University of California San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093, USA.

tumour-promoting machinery. This suggestion is based on the realization that NF- κ B is a major activator of anti-apoptotic gene expression^{24–27}. These findings were first made in the context of TNF- α signalling (Fig. 3). Despite the presence of death domains (DDs) in the intracellular portion of its major receptor TNFR1, TNF- α does not trigger apoptosis unless it is combined with inhibitors of RNA or protein synthesis. The requirement for such inhibitors can be alleviated through inactivation of NF- κ B by either deletion of its RelA subunit or expression of a degradation-resistant form of inhibitor of NF- κ B (I κ B) — the so-called I κ B-superrepressor (I κ B-SR). A similar effect is seen upon deletion of *Ikkkb* (I κ B kinase (IKK) β -subunit) or *Ikkkg* (IKK γ -subunit)²⁸. Such interventions were found to inhibit the expression of several critical anti-apoptotic proteins, including the specific inhibitor of caspase 8 activation c-FLIP²⁹, the caspase inhibitors cIAP1 and cIAP2, and the anti-apoptotic member of the B-cell leukaemia/lymphoma 2 (Bcl2) family Bcl-X_L (ref. 30). As it was already established that genes encoding anti-apoptotic proteins — for instance Bcl2 — function as oncogenes³¹, it was reasonable to predict that NF- κ B activation in chronic infection and inflammation could also promote tumour development¹¹. Although this hypothesis was consistent with the presence of activated NF- κ B in many cancers, it required validation in appropriate animal models. Because the deletion of individual NF- κ B subunit genes, with the exception of RelA, resulted in partially redundant phenotypes, two complementary approaches were undertaken to inhibit most forms of NF- κ B in the mouse, and to examine its role in tumorigenesis. To avoid the embryonic lethality associated with the loss of a substantial amount of NF- κ B activity, both approaches were based on the cell type-specific inhibition of NF- κ B, in one case by expressing I κ B-SR from a cell-specific promoter³², and in the other case through conditional inactivation of the *Ikkkb* gene³³. Both methods resulted in substantial inhibition of the classical NF- κ B signalling pathway, with no direct effect on the alternative pathway (Box 1). The results of these mouse studies are described below.

NF- κ B in inflammation-linked cancers

Colitis-associated cancer (CAC) is a colorectal disease that arises in patients suffering from the chronic IBD UC. In mice, CAC can be modelled by injection of the procarcinogen azoxymethane (AOM), which undergoes metabolic activation in intestinal epithelial cells (enterocytes). AOM exposure alone causes cancer with low incidence, but this can be augmented by three rounds of exposure to dextran-sulphate sodium salt (DSS), which is an irritant that causes colonic inflammation

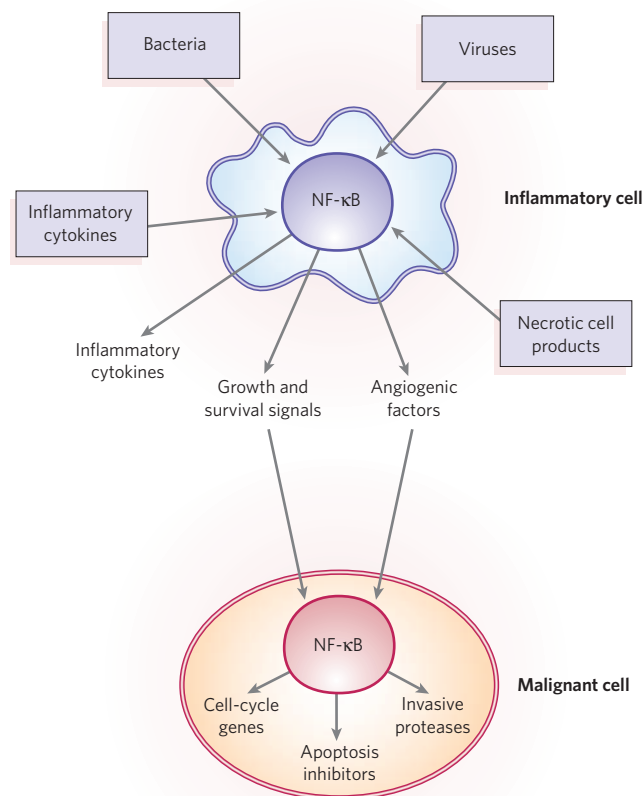


Figure 1 | NF- κ B activation, and the interaction between inflammatory and malignant cells, can promote malignant conversion and progression. Activation of nuclear factor- κ B (NF- κ B) in inflammatory cells in response to infectious agents, inflammatory cytokines and proteins, and danger signals released by necrotic cells lead to the production of secreted factors that enhance the growth, survival and vascularization of carcinoma cells. Activation of NF- κ B in the latter results in elevated expression of cell-cycle genes (such as cyclin D1), inhibitors of apoptosis (such as B-cell leukaemia/lymphoma-X_L) and proteases that promote the invasive phenotype (such as matrix metalloproteinase-9).

Box 1 | Nuclear factor- κ B transcription factors and their regulation

Nuclear factor- κ B (NF- κ B) transcription factors are assembled through the dimerization of five subunits: RelA (p65), c-Rel, RelB, p50/NF- κ B1 and p52/NF- κ B2 (ref. 59). The first clue linking NF- κ B to cancer was the recognition that *c-rel*, which is the cellular homologue of the *v-rel* oncogene, encodes an NF- κ B subunit, and that all of these proteins share the same DNA binding domain — the Rel homology domain⁶⁰. Not surprisingly, overexpression of normal Rel proteins promotes oncogenic transformation.

Before cell stimulation, most NF- κ B dimers are retained in the cytoplasm by binding to specific inhibitors — the inhibitors of NF- κ Bs (I κ Bs). Cell stimulation activates the I κ B kinase (IKK) complex, which is composed of two catalytic subunits (IKK- α and IKK- β) and a regulatory subunit (IKK- γ /NEMO)⁶¹. Activated IKK phosphorylates NF- κ B-bound I κ B proteins, and targets them for polyubiquitination and rapid degradation by creating a binding site for the SCF^{TRCP} ubiquitin ligase complex⁶². Freed NF- κ B dimers translocate to the nucleus where they coordinate the transcriptional activation of several hundred target genes, many of which also depend on other transcription factors^{63–65}.

Although it is of fundamental importance, this pathway, which is called the classical NF- κ B signalling pathway, is only one of two major pathways that activate NF- κ B transcription factors. The second pathway, the alternative pathway, results in specific activation of p52:RelB heterodimers and is not required for activation of the more ubiquitous

p50:RelA dimers⁶⁶. Unlike the classical pathway, which is dependent on IKK- γ and to a large extent on the activity of IKK- β , the alternative pathway is based on IKK- α homodimers, the preferred substrate of which is the precursor of p52 — p100/NF- κ B2 (ref. 67). This protein binds RelB through its amino-terminal Rel homology domain, and keeps itself and its partner in the cytoplasm through its carboxy-terminal I κ B-like domain. Activation of IKK- α dimers results in degradation of the latter and nuclear entry of p52:RelB dimers.

Although this pathway is not directly involved in innate immunity and inflammation, it is required for the generation of secondary lymphoid organs, and for B-cell maturation and survival⁶⁶. Its relevance to cancer is underscored by chromosomal translocations associated with B-cell lymphoma that remove the region encoding the I κ B-like inhibitory domain of p100 (ref. 68). In addition, the alternative pathway might be involved in mammary carcinogenesis⁶⁹. Mutations that affect other NF- κ B subunits were identified in other types of cancer, especially those of lymphoid origin¹².

Another pathway that can lead to NF- κ B activation is independent of IKK and, instead, is based on activation of casein kinase 2 (CK2), which induces I κ B α degradation through the phosphorylation of carboxy-terminal sites⁷⁰. This pathway has only a minor role in physiological NF- κ B activation, although it might contribute to skin carcinogenesis because it is activated by ultraviolet radiation.

(colitis) by eroding the mucosal barrier, thereby exposing lamina propria macrophages to normal enteric bacteria³⁴. These macrophages are activated to produce a range of inflammatory mediators. Selective inactivation of the *Ikkbb* gene within enterocytes resulted in an 80% decrease in CAC tumour multiplicity³⁵. As tumour size was not affected, it can be concluded that IKK- β -dependent NF- κ B in enterocytes contributes to tumour initiation or early tumour promotion, rather than tumour growth and progression. Indeed, analysis of enterocyte IKK- β -deleted mice shortly after exposure to AOM plus DSS revealed increased apoptosis of enterocytes, including pre-neoplastic cells in which AOM led to mutational activation of the β -catenin pathway³⁵. Enhanced apoptosis is probably caused by defective induction of Bcl-X_L. However, when IKK- β was deleted in myeloid cells (for example, mature macrophages, dendritic cells and neutrophils), tumour multiplicity was reduced by only 50%, although tumour size was also reduced³⁵. Indeed, deletion of IKK- β in myeloid cells, but not in enterocytes, diminished the proliferation of AOM-exposed enterocytes. The myeloid-specific mutation, however, had no effect on apoptosis of AOM-exposed enterocytes. These results led to the conclusion that IKK- β -driven NF- κ B contributes to the development of CAC through two distinct cell-type-specific mechanisms: in enterocytes it activates anti-apoptotic genes and thereby suppresses the apoptotic elimination of pre-neoplastic cells, whereas in myeloid cells it promotes the production of cytokines that act as growth factors for pre-malignant enterocytes. One of these growth factors was subsequently identified as IL-6, which is encoded by an NF- κ B target gene³⁶. The inhibition of IL-6 signalling with antagonistic anti-IL-6 receptor antibodies inhibited tumour growth with little effect on tumour multiplicity, thereby resembling IKK- β ablation in myeloid cells. Curiously, in the early stages of the carcinogenic protocol, IL-6 is produced by lamina propria myeloid cells³⁵, whereas at the end of the CAC protocol it is mainly expressed by tumour-infiltrating T cells³⁶. As with other chronic inflammatory responses¹⁵, the colonic inflammation induced by DSS seems to be initiated through rapid activation of tissue macrophages but is sustained through prolonged activation of pro-inflammatory T cells.

Another interesting model of inflammation-driven cancer is the multidrug resistance 2 (*Mdr2*)-knockout mouse, in which the absence of the MDR2 transporter leads to accumulation of bile acids and phospholipids within hepatocytes, resulting in low-grade liver inflammation, which eventually gives rise to HCC at 8–10 months of age³⁷. Unlike the CAC model described above, the initiating event that leads to cancer development in *Mdr2*^{-/-} mice is not known, because inflammation probably only accounts for tumour promotion. Nonetheless, as in the CAC model, inhibition of NF- κ B through expression of I κ B-SR under the control of a promoter that is highly active in hepatocytes blocked tumour development²¹. Exactly how bile acid and phospholipid accumulation lead to NF- κ B activation in this model is not clear, but it seems to depend on the production of TNF- α by non-parenchymal cells (that is, Kupffer and endothelial cells). Indeed, inhibition of TNF- α signalling prevented activation of NF- κ B in hepatocytes and early tumours, and, just like the inhibition of NF- κ B itself, increased the number of apoptotic hepatocytes and reduced tumour multiplicity²¹. Thus, in this model too, an important pro-tumorigenic function of NF- κ B is the suppression of apoptosis of pre-malignant or early neoplastic cells (Fig. 1). The role of NF- κ B in inflammatory cells has not been investigated in this model but is presumably important for the production of TNF- α and other cytokines.

Additional inflammation-driven cancers include lymphomas of mucosal-associated lymphoid tissue (MALT), which develop in the context of prolonged lymphoid proliferation caused by chronic microbial infections, such as *H. pylori* gastritis³⁸. Interestingly, MALT lymphoma is associated with elevated expression of proteins that are involved in IKK and NF- κ B activation in response to the occupancy of antigen receptors, namely Bcl10 (ref. 39) and MALT1 (refs 40, 41). Presumably, MALT lymphoma is initiated by repeated antigenic stimulation of B cells and, up to a certain point, can be reversed by antibiotic treatment. Later on, however, chromosomal translocations, which are more frequent in

activated B cells, place the *Bcl10* and *MALT1* genes under the control of heterologous promoters and cause their overexpression, most commonly as fusion proteins (for example, IAP2–MALT1), thereby leading to persistent activation of NF- κ B. This suppresses B-cell apoptosis and contributes to uncontrolled B-cell proliferation. At this point, the tumours no longer regress upon elimination of the antigenic challenge. Although Bcl10-deficient and MALT1-deficient mice have been generated, and used to confirm the role of these molecules in NF- κ B activation^{39,41}, an appropriate mouse model of MALT lymphoma is yet to be described. Nonetheless, there is little doubt that MALT lymphoma is caused by the constitutive activation of NF- κ B.

NF- κ B and chemically induced liver and skin cancers

Although NF- κ B has emerged as a critical promoter of inflammation-linked cancers, there is also strong evidence that it has the opposite effect in models of chemically induced skin and liver cancers. The first evidence for such an effect came from a two-stage skin carcinogenesis model based on 7,12-dimethylbenz(a)anthracene (DMBA) as a tumour initiator and the phorbol ester TPA as a tumour promoter. Inhibition of NF- κ B in keratinocytes greatly enhanced the multiplicity of squamous cell carcinomas (SCCs) caused by exposure to DMBA plus TPA^{42,43}. Similarly, inhibition of NF- κ B in primary human keratinocytes promoted their Ras-mediated transformation⁴⁴, indicating that the anti-tumorigenic effect is not limited to the two-stage model. The tumour-suppressing activity of NF- κ B was explained by either inhibition of cell-cycle progression^{43,44} or downmodulation of Jun kinase (JNK) activity⁴⁵. JNK is a member of the MAP kinase (MAPK) family that stimulates activator protein-1 (AP-1) transcription factors⁴⁶. It is well established that NF- κ B activation in response to TNFR1 engagement (but not other

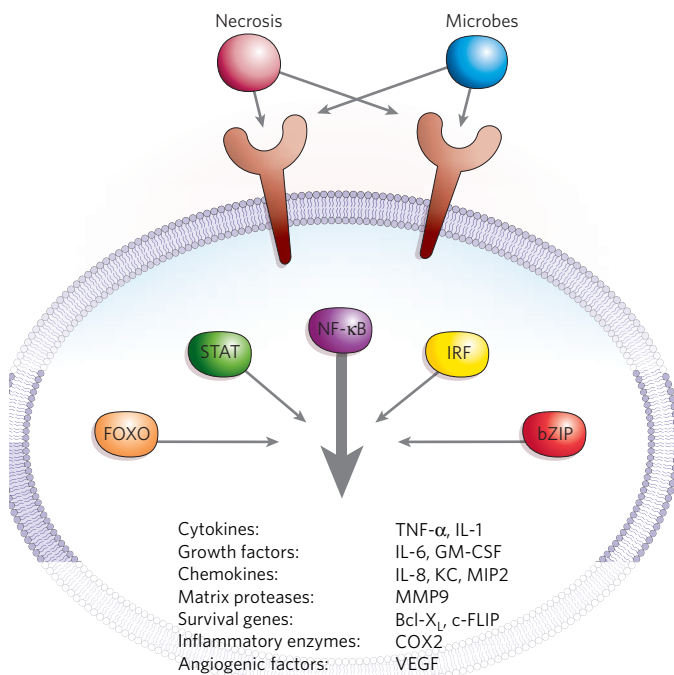


Figure 2 | Microbial pathogens and tissue necrosis lead to activation of NF- κ B and other transcription factors in cells that express pattern-recognition receptors. Activation of NF- κ B and other transcription factors involved in the innate immune/inflammatory response can upregulate the expression of many genes, the products of which promote and support the malignant phenotype. Bcl-X_L, B-cell leukaemia/lymphoma-X_L; bZIP, basic leucine zipper protein; c-FLIP, c-FLICE-inhibitory protein; COX-2, cyclooxygenase-2; FOXO, forkhead transcription factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; IL, interleukin; IRF, interferon response factor; KC, Cxcl1 chemokine; MIP2, macrophage inflammatory protein-2; MMP9, matrix metalloproteinase-9; STAT, signal transducer and activator of transcription; TNF- α , tumour-necrosis factor- α ; VEGF, vascular endothelial growth factor.

The negative interplay between NF- κ B and JNK was also found to have a critical role in the development of chemically induced HCC in response to administration of the pro-carcinogen diethylnitrosamine (DEN). Unlike the carcinogens used in CAC or two-stage skin carcinogenesis, DEN does not require any assistance from inflammation-inducing tumour promoters if it is given to 2-week-old male mice. In contrast to the *Mdr2*^{-/-} HCC model, but similar to two-stage skin carcinogenesis, hepatocyte-specific IKK- β ablation greatly augmented HCC multiplicity and size in DEN-treated mice⁴⁹. Curiously, DEN administration induces rapid TNF- α production and activation of TNFR1 (T. Sakurai and M.K., unpublished observations); thus, TNFR1 signalling might also be important in this cancer model. Decreased NF- κ B activity and elevated JNK activity promote TNF- α -induced cell death^{47,50} (Fig. 3); accordingly, ablation of hepatocyte IKK- β results in higher DEN-induced JNK activity and more cell death⁴⁹. However, owing to the strong regenerative capacity of the liver, elevated hepatocyte death enhances compensatory proliferation. Prolonged JNK activation in the absence of NF- κ B depends on the accumulation of reactive oxygen species⁴⁷ (Fig. 3). Correspondingly, feeding hepatocyte-IKK- β -deficient mice with the potent anti-oxidant butylated hydroxyanisole (BHA) prevented DEN-induced prolonged JNK activation and excessive hepatocyte death, resulting in strong inhibition of compensatory proliferation⁴⁹. Furthermore, feeding BHA before DEN administration⁴⁹, or ablation of the major death-promoting and growth-promoting JNK isoform JNK1 (T. Sakurai and M.K., unpublished observations), prevented the increase in hepatocarcinogenesis seen in hepatocyte-IKK- β -deficient mice. Notably, the increase in hepatocyte proliferation in the absence of IKK- β is not due solely to direct effects of NF- κ B or JNK1 on the hepatocyte cell-cycle machinery. Much of the compensatory proliferation that is observed depends on the production of factors such as TNF- α , IL-6 and hepatocyte growth factor (HGF) by non-parenchymal cells. Correspondingly, ablation of IKK- β in liver myeloid cells, which are known as Kupffer cells, prevented the induction of these cytokines in response to DEN administration, thereby resulting in a marked decrease in HCC load⁴⁹. The production of IL-6 by Kupffer cells might depend on their activation by proteins that are released by necrotic hepatocytes, which are potent macrophage activators at least *in vitro*^{49,51}. The induction of TNF- α production by DEN is another likely contributor to this carcinogenesis model, as found in the two-stage skin carcinogenesis model^{19,20}, and could be the reason why inhibition of NF- κ B potentiates carcinogenesis in both of these models.

Although NF- κ B can either promote or oppose tumour development, several general mechanisms of action emerge from the work discussed above. First and foremost, NF- κ B is an activator of anti-apoptotic genes³⁰. Both in CAC and cholestatic liver cancer, the activation of NF- κ B within pre-neoplastic or progressing cancer cells is the rate-limiting event, which ensures their survival and prevents elimination by pro-apoptotic tumour-surveillance mechanisms. However, as a result of different tissue kinetics, inhibition of cell death — as seen in hepatocytes exposed to the cytotoxic chemical DEN — prevents compensatory proliferation, thereby attenuating tumour development. Compensatory proliferation is critical for DEN-induced hepatocarcinogenesis because the target population for this carcinogen comprises non-proliferating differentiated hepatocytes⁴⁹. In contrast to the liver, the colonic epithelium is subject to constant renewal, and compensatory proliferation is not a rate-limiting event in CAC.

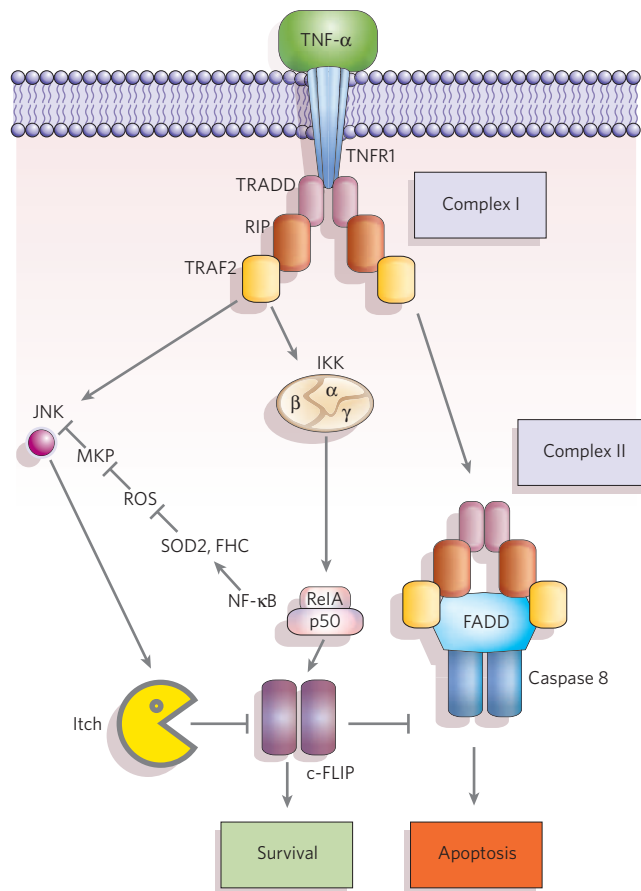


Figure 3 | TNFR1 signalling controls cell survival and death by through an interplay between NF- κ B and JNK. Activation of tumour-necrosis factor receptor 1 (TNFR1) by binding of TNF- α results in rapid assembly of complex I, which is composed of TNF receptor 1-associated protein (TRADD), receptor-interacting protein 1 (RIP1) and TNFR-associated factor 2/5 (TRAF2/5). This complex leads to activation of inhibitor of NF- κ B kinase (IKK) and the Jun kinase (JNK) cascade. Activation of IKK leads to nuclear translocation of NF- κ B and upregulation of several anti-apoptotic genes, including the gene encoding c-FLICE-inhibitory protein (c-FLIP), which is a specific inhibitor of caspase 8 activation. NF- κ B also upregulates the expression of genes encoding the antioxidant proteins superoxide dismutase (SOD2) and ferritin heavy chain (FHC), which block the accumulation of reactive oxygen species (ROS) and prevent the inhibition of MAP kinase phosphatases (MKPs). Overall, this promotes the rapid termination of JNK activity and, generally, signalling through complex I enhances cell survival. After some time, complex I dissociates and the soluble complex II is formed, which, in addition to TRADD, RIP1 and TRAF2/5, includes the adaptor protein Fas-associated death domain protein (FADD). The latter can recruit and activate caspase 8, and thereby trigger apoptosis, but this does not occur if c-FLIP levels are high. Defective activation of NF- κ B prevents *de novo* synthesis of c-FLIP, resulting in accumulation of ROS and inhibition of MKPs, thereby promoting prolonged JNK activation. This leads to extended phosphorylation and activation of the ubiquitin ligase Itch, which promotes c-FLIP degradation. Once the level of c-FLIP falls below a certain threshold, caspase 8 is activated and the cell undergoes apoptosis.

Importantly, even in those cases in which NF- κ B activation in epithelial cells negatively affects tumour development, its activation in inflammatory cells has the opposite effect. In the four experimental cancer models mentioned above, tumour promotion and progression largely depend on production of the proinflammatory cytokines TNF- α and IL-6, which serve as growth factors for pre-malignant cells and already formed tumours. Interference with either TNF- α or IL-6 signalling blocks tumour growth and progression in the *Mdr2*^{-/-} and CAC models^{21,36}. Even in SCC, the development of which is enhanced by inhibition of NF- κ B, interference with TNF- α signalling blocks tumour

development²⁰. Other NF- κ B target genes encode chemokines such as the mouse orthologues of human IL-8: MIP2, KC and Gro1. Interference with Gro1 signalling inhibits the development of colorectal cancer in mice⁵², whereas IL-8 (CXCL8) inhibition blocks the growth of a Ras-transformed human xenograft in these animals⁵³. Inhibition of IL-8 produced by malignant cells prevents tumour angiogenesis by blocking inflammatory cell recruitment. Additional evidence for a role of NF- κ B in tumour angiogenesis comes from studies on the *ING4* tumour-suppressor gene, the expression of which is reduced in gliomas. *ING4* directly interacts with RelA and inhibits expression of angiogenesis-promoting genes, such as *IL8* and cyclooxygenase 2 (*COX2*), by interfering with NF- κ B-dependent transcription⁵⁴.

In addition to effects on cell survival, and the production of growth and angiogenesis factors, NF- κ B might directly stimulate cell-cycle progression through transcriptional activation of cell-cycle genes, especially cyclin D1 (ref. 55). Interestingly, in mammary epithelial cells⁵⁶ and ErbB2-transformed mammary carcinomas (Y. Cao and M.K., unpublished observations), expression of cyclin D1 depends on IKK- α ; however, this function of IKK- α is mediated through the classical, and not the alternative, pathway. NF- κ B is also involved in TNF- α -mediated induction of cyclin D1 and other cell-cycle genes in mouse colorectal carcinoma⁵⁷. Yet, in epidermal keratinocytes, NF- κ B is a negative regulator of cell proliferation⁴⁴. This function of NF- κ B was suggested to be mediated through suppression of JNK activation⁴⁵. Importantly, JNK activation is also an important contributor to chemically induced HCC⁴⁹. Thus, in addition to being an important regulator of cell survival in the context of TNF- α signalling (Fig. 3), the NF- κ B–JNK interplay seems to be a critical regulator of cancer development and progression, especially in cases where tissue injury triggers compensatory proliferation. Thus, the role of epithelial cell NF- κ B in tumorigenesis is highly dependent on its overall effect on tissue kinetics.

NF- κ B as a target for cancer prevention and therapy

Given the drastic and invariable effects of inflammatory cell NF- κ B on tumour development and progression, this factor and the signalling pathways involved in its activation are attractive targets for cancer prevention and therapy. Although epithelial-specific inhibition of NF- κ B might increase cancers of the liver and skin, it should be recognized that clinically relevant inhibitors of this pathway are unlikely to act in a cell-type-specific manner, and, as long as NF- κ B is inhibited in inflammatory cells, the net effect should be less cancer and slower tumour development. However, NF- κ B in inflammatory cells serves an important immune function, and its absence can result in severe immunodeficiency. Thus, prolonged and substantial inhibition of NF- κ B might not be practical in cancer prevention. More effective preventive measures might be those that are directed at the initial causes of persistent NF- κ B activation: microbial and viral infections or chronic inflammatory disorders. However, the prolonged use of any anti-inflammatory drug, regardless of its impact on NF- κ B, is likely to have adverse side effects that cannot be tolerated in cancer prevention.

Given these considerations, NF- κ B inhibitors are more likely to be of use in cancer therapy, in which they can be administered intermittently for shorter durations, thereby avoiding immunosuppression associated with long-term inhibition. Although, in some cases, NF- κ B inhibition contributes to tumour development, the most likely outcome of its inhibition in existing tumours is increased cancer cell apoptosis. Nonetheless, in most cases, the mere inhibition of NF- κ B is insufficient for a pronounced apoptotic response unless combined with apoptosis-inducing drugs or radiation. Thus, NF- κ B inhibitors are most likely to be used as adjuvants along with other cancer therapies. As recently reviewed⁵⁸, several IKK and NF- κ B inhibitors are under development, as well as a number of natural products that can inhibit NF- κ B activation when used at high doses but can also affect several other targets. There are no published accounts, as yet, on the use of specific NF- κ B inhibitors in cancer therapy, but given the great interest in the role of NF- κ B in cancer development and progression, and the current emphasis on clinical translation of basic research findings, this situation is likely to

change in the not so distant future. Progress in this area will, of course, depend not only on the development of potent and specific, orally available IKK and NF- κ B inhibitors, but also on the direct demonstration of an involvement of the specific pathways discussed above in various human malignancies.

- Balkwill, F. & Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet* **357**, 539–545 (2001).
- Coussens, L. M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
- Kuper, H., Adami, H. O. & Trichopoulos, D. Infections as a major preventable cause of human cancer. *J. Intern. Med.* **248**, 171–183 (2000).
- Roder, D. M. The epidemiology of gastric cancer. *Gastric Cancer* **5** (Suppl. 1), 5–11 (2002).
- Ekbom, A. Risk of cancer in ulcerative colitis. *J. Gastrointest. Surg.* **2**, 312–313 (1998).
- Dhala, A., Pinsker, K. & Prezant, D. J. Respiratory health consequences of environmental tobacco smoke. *Med. Clin. North Am.* **88**, 1535–1552, xi (2004).
- Greten, F. R. & Karin, M. NF- κ B: linking inflammation and immunity to cancer development and progression. *Nature Rev. Immunol.* **5**, 749–759 (2005).
- Balkwill, F., Charles, K. A. & Mantovani, A. Smoldering and polarized inflammation in the initiation and promotion of malignant disease. *Cancer Cell* **7**, 211–217 (2005).
- Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
- Li, Q., Withoff, S. & Verma, I. M. Inflammation-associated cancer: NF- κ B is the lynchpin. *Trends Immunol.* **26**, 318–325 (2005).
- Karin, M., Cao, Y., Greten, F. R. & Li, Z. W. NF- κ B in cancer: from innocent bystander to major culprit. *Nature Rev. Cancer* **2**, 301–310 (2002).
- Gilmore, T. D. The Rel/NF- κ B/I κ B signal transduction pathway and cancer. *Cancer Treat. Res.* **115**, 241–265 (2003).
- Medzhitov, R. Toll-like receptors and innate immunity. *Nature Rev. Immunol.* **1**, 135–145 (2001).
- Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
- Nizet, V., Lawrence, T. & Karin, M. Innate immunity gone awry: mechanisms linking microbial infections to chronic inflammatory disorders and cancer. *Cell* **124**, 823–835 (2006).
- Li, Q. & Verma, I. M. NF- κ B regulation in the immune system. *Nature Rev. Immunol.* **2**, 725–734 (2002).
- Sun, J. et al. Sequence variants in Toll-like receptor gene cluster (*TLR6–TLR1–TLR10*) and prostate cancer risk. *J. Natl Cancer Inst.* **97**, 525–532 (2005).
- El-Omar, E. M. et al. Interleukin-1 polymorphisms associated with increased risk of gastric cancer. *Nature* **404**, 398–402 (2000).
- Arnott, C. H. et al. Tumour-necrosis factor- α mediates tumour promotion via a PKC α - and AP-1-dependent pathway. *Oncogene* **21**, 4728–4738 (2002).
- Lind, M. H. et al. Tumor necrosis factor receptor 1-mediated signaling is required for skin cancer development induced by NF- κ B inhibition. *Proc. Natl Acad. Sci. USA* **101**, 4972–4977 (2004).
- Pikarsky, E. et al. NF- κ B functions as a tumour promoter in inflammation-associated cancer. *Nature* **431**, 461–466 (2004).
- Balkwill, F. Tumor necrosis factor or tumor promoting factor? *Cytokine Growth Factor Rev.* **13**, 135–141 (2002).
- Oshima, M., Oshima, H., Matsunaga, A. & Taketo, M. M. Hyperplastic gastric tumors with spasmodic polypeptide-expressing metaplasia caused by tumor necrosis factor- α -dependent inflammation in cyclooxygenase-2/microsomal prostaglandin E synthase-1 transgenic mice. *Cancer Res.* **65**, 9147–9151 (2005).
- Liu, Z.-G., Hu, H., Goeddel, D. V. & Karin, M. Dissection of TNF receptor 1 effector functions: JNK activation is not linked to apoptosis, while NF- κ B activation prevents cell death. *Cell* **87**, 565–576 (1996).
- Beg, A. A. & Baltimore, D. An essential role for NF- κ B in preventing TNF- α induced cell death. *Science* **274**, 782–784 (1996).
- Van Antwerp, D. J., Martin, S. J., Kafri, T., Green, D. R. & Verma, I. M. Suppression of TNF α -induced apoptosis by NF- κ B. *Science* **274**, 787–789 (1996).
- Wang, C.-Y., Mayo, M. W. & Baldwin, A. S. Jr. TNF- α and cancer therapy-induced apoptosis: potentiation by inhibition of NF- κ B. *Science* **274**, 784–787 (1996).
- Makris, C. et al. Female mice heterozygous for IKK γ /NEMO deficiencies develop a dermatopathy similar to the human X-linked disorder incontinentia pigmenti. *Mol. Cell* **5**, 969–979 (2000).
- Micheau, O., Lens, S., Gaide, O., Alevizopoulos, K. & Tschopp, J. NF- κ B signals induce the expression of c-FLIP. *Mol. Cell Biol.* **21**, 5299–5305 (2001).
- Karin, M. & Lin, A. NF- κ B at the crossroads of life and death. *Nature Immunol.* **3**, 221–227 (2002).
- Danial, N. N. & Korsmeyer, S. J. Cell death: critical control points. *Cell* **116**, 205–219 (2004).
- Lavon, I. et al. High susceptibility to bacterial infection, but no liver dysfunction, in mice compromised for hepatocyte NF- κ B activation. *Nature Med.* **6**, 573–577 (2000).
- Chen, L. W. et al. The two faces of IKK and NF- κ B inhibition: prevention of systemic inflammation but increased local injury following intestinal ischemia-reperfusion. *Nature Med.* **9**, 575–581 (2003).
- Okayasu, I., Ohkusa, T., Kajura, K., Kanno, J. & Sakamoto, S. Promotion of colorectal neoplasia in experimental murine ulcerative colitis. *Gut* **39**, 87–92 (1996).
- Greten, F. R. et al. IKK β links inflammation and tumorigenesis in a mouse model of colitis-associated cancer. *Cell* **118**, 285–296 (2004).
- Becker, C. et al. TGF- β suppresses tumor progression in colon cancer by inhibition of IL-6 trans-signaling. *Immunity* **21**, 491–501 (2004).
- Mauad, T. H. et al. Mice with homozygous disruption of the *mdr2* P-glycoprotein gene. A novel animal model for studies of nonsuppurative inflammatory cholangitis and hepatocarcinogenesis. *Am. J. Pathol.* **145**, 1237–1245 (1994).
- Zucca, E., Bertoni, F., Roggero, E. & Cavalli, F. The gastric marginal zone B-cell lymphoma of MALT type. *Blood* **96**, 410–419 (2000).

39. Ruland, J. *et al.* Bcl10 is a positive regulator of antigen receptor-induced activation of NF- κ B and neural tube closure. *Cell* **104**, 33–42 (2001).
40. Lucas, P. C. *et al.* Bcl10 and MALT1, independent targets of chromosomal translocation in malt lymphoma, cooperate in a novel NF- κ B signaling pathway. *J. Biol. Chem.* **276**, 19012–19019 (2001).
41. Thome, M. CARMA1, BCL-10 and MALT1 in lymphocyte development and activation. *Nature Rev. Immunol.* **4**, 348–359 (2004).
42. van Hogerlinden, M., Rozell, B. L., Ahrlund-Richter, L. & Toftgard, R. Squamous cell carcinomas and increased apoptosis in skin with inhibited Rel/nuclear factor- κ B signaling. *Cancer Res.* **59**, 3299–3303 (1999).
43. van Hogerlinden, M., Auer, G. & Toftgard, R. Inhibition of Rel/nuclear factor- κ B signaling in skin results in defective DNA damage-induced cell cycle arrest and Ha-ras- and p53-independent tumor development. *Oncogene* **21**, 4969–4977 (2002).
44. Dajee, M. *et al.* NF- κ B blockade and oncogenic Ras trigger invasive human epidermal neoplasia. *Nature* **421**, 639–643 (2003).
45. Zhang, J. Y., Green, C. L., Tao, S. & Khavari, P. A. NF- κ B RelA opposes epidermal proliferation driven by TNFR1 and JNK. *Genes Dev.* **18**, 17–22 (2004).
46. Karin, M. The regulation of AP-1 activity by mitogen-activated protein kinases. *J. Biol. Chem.* **270**, 16483–16486 (1995).
47. Kamata, H. *et al.* Reactive oxygen species promote TNF α -induced death and sustained JNK activation by inhibiting MAP kinase phosphatases. *Cell* **120**, 649–661 (2005).
48. Pham, C. G. *et al.* Ferritin heavy chain upregulation by NF- κ B inhibits TNF α -induced apoptosis by suppressing reactive oxygen species. *Cell* **119**, 529–542 (2004).
49. Maeda, S., Kamata, H., Luo, J. L., Leffert, H. & Karin, M. IKK β couples hepatocyte death to cytokine-driven compensatory proliferation that promotes chemical hepatocarcinogenesis. *Cell* **121**, 977–990 (2005).
50. Chang, L. *et al.* The E3 ubiquitin ligase itch couples JNK activation to TNF α -induced cell death by inducing c-FLIP(L) turnover. *Cell* **124**, 601–613 (2006).
51. Park, J. S. *et al.* Involvement of toll-like receptors 2 and 4 in cellular activation by high mobility group box 1 protein. *J. Biol. Chem.* **279**, 7370–7377 (2004).
52. Holla, V. R. *et al.* Prostaglandin E2 regulates the complement inhibitor CD55/decay-accelerating factor in colorectal cancer. *J. Biol. Chem.* **280**, 476–483 (2005).
53. Sparmann, A. & Bar-Sagi, D. Ras-induced interleukin-8 expression plays a critical role in tumor growth and angiogenesis. *Cancer Cell* **6**, 447–458 (2004).
54. Garkavtsev, I. *et al.* The candidate tumour suppressor protein ING4 regulates brain tumour growth and angiogenesis. *Nature* **428**, 328–332 (2004).
55. Joyce, D. *et al.* NF- κ B and cell-cycle regulation: the cyclin connection. *Cytokine Growth Factor Rev.* **12**, 73–90 (2001).
56. Cao, Y. *et al.* IKK α provides an essential link between RANK signaling and cyclin D1 expression during mammary gland development. *Cell* **107**, 763–775 (2001).
57. Luo, J. L., Maeda, S., Hsu, L. C., Yagita, H. & Karin, M. Inhibition of NF- κ B in cancer cells converts inflammation-induced tumor growth mediated by TNF α to TRAIL-mediated tumor regression. *Cancer Cell* **6**, 297–305 (2004).
58. Karin, M., Yamamoto, Y. & Wang, Q. M. The IKK NF- κ B system: a treasure trove for drug development. *Nature Rev. Drug Discov.* **3**, 17–26 (2004).
59. Ghosh, S. & Karin, M. Missing pieces in the NF- κ B puzzle. *Cell* **109** (suppl.), S81–S96 (2002).
60. Gilmore, T. D. The Rel/NF- κ B signal transduction pathway: introduction. *Oncogene* **18**, 6842–6844 (1999).
61. Rothwarf, D. M. & Karin, M. The NF- κ B activation pathway: a paradigm in information transfer from membrane to nucleus. *Sci. STKE* **1999**, RE1 (1999).
62. Karin, M. & Ben-Neriah, Y. Phosphorylation meets ubiquitination: the control of NF- κ B activity. *Annu. Rev. Immunol.* **18**, 621–663 (2000).
63. Werner, S. L., Barken, D. & Hoffmann, A. Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science* **309**, 1857–1861 (2005).
64. Park, J. M. *et al.* Signaling pathways and genes that inhibit pathogen-induced macrophage apoptosis — CREB and NF- κ B as key regulators. *Immunity* **23**, 319–329 (2005).
65. Covert, M. W., Leung, T. H., Gaston, J. E. & Baltimore, D. Achieving stability of lipopolysaccharide-induced NF- κ B activation. *Science* **309**, 1854–1857 (2005).
66. Bonizzi, G. & Karin, M. The two NF- κ B activation pathways and their role in innate and adaptive immunity. *Trends Immunol.* **25**, 280–288 (2004).
67. Sentfleben, U. *et al.* Activation by IKK α of a second, evolutionary conserved, NF- κ B signaling pathway. *Science* **293**, 1495–1499 (2001).
68. Neri, A. *et al.* B-cell lymphoma-associated chromosomal translocation involves candidate oncogene *lyt-10*, homologous to NF- κ B p50. *Cell* **67**, 1075–1087 (1991).
69. Demicco, E. G. *et al.* RelB/p52 NF- κ B complexes rescue an early delay in mammary gland development in transgenic mice with targeted superrepressor I κ B- α expression and promote carcinogenesis of the mammary gland. *Mol. Cell. Biol.* **25**, 10136–10147 (2005).
70. Kato, T. Jr., Delhase, M., Hoffmann, A. & Karin, M. CK2 is a c-terminal I κ B kinase responsible for NF- κ B activation during the UV response. *Mol. Cell* **12**, 829–839 (2003).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to M.K. (karinoffice@ucsd.edu).

Hypoxia signalling in cancer and approaches to enforce tumour regression

Jacques Pouyssegur¹, Frédéric Dayan¹ & Nathalie M. Mazure¹

Tumour cells emerge as a result of genetic alteration of signal circuitries promoting cell growth and survival, whereas their expansion relies on nutrient supply. Oxygen limitation is central in controlling neovascularization, glucose metabolism, survival and tumour spread. This pleiotropic action is orchestrated by hypoxia-inducible factor (HIF), which is a master transcriptional factor in nutrient stress signalling. Understanding the role of HIF in intracellular pH (pH_i) regulation, metabolism, cell invasion, autophagy and cell death is crucial for developing novel anticancer therapies. There are new approaches to enforce necrotic cell death and tumour regression by targeting tumour metabolism and pH_i-control systems.

We have learned, over the past two decades, how mammalian cells perceive signals to induce cell-cycle progression, proliferation and survival. Two major pathways that are frequently mutated in human cancer, the Ras–extracellular signal-regulated kinase (ERK)^{1–3} and the phosphatidylinositol-3-OH kinase (PI(3)K)–AKT⁴ (see the review in this issue by Shaw and Cantley, page 424) signalling cascades, are activated by a vast array of growth factor polypeptides, hormones and extracellular matrix proteins⁵. Activation of these two pathways is sufficient to trigger multiple cycles of division and survival of normal cells under the ‘rich’ conditions of tissue culture. *In vivo*, however, growing cells must constantly instruct the microenvironment to maintain a supply of essential nutrients. It is remarkable that the Ras–ERK and PI(3)K–AKT pathways also control the expression of the ubiquitous vascular endothelial growth factor-A (VEGF-A), which is a key factor in vascularization/angiogenesis^{6,7}. During embryonic development or in the context of tumour expansion, growing cells rapidly outstrip the supply of nutrients. Although cells sense and respond to variations in concentrations of all nutrients, oxygen sensing has emerged as a central control mechanism of vasculogenesis^{8,9}.

At the heart of this regulatory system is HIF^{10,11}, which controls, among other gene products, the expression of two key angiogenic factors: VEGF-A¹² and angiopoietin-2 (Ang-2)¹³. This finding has placed the hypoxia-signalling pathway at the forefront of nutritional control — a notion reinforced by the fact that growth factors enhance HIF expression and converge with hypoxia in inducing maximal expression of VEGF-A.

HIF can induce a vast array of gene products controlling energy metabolism, neovascularization, survival, pH_i and cell migration, and has become recognized as a strong promoter of tumour growth¹⁴. This pro-oncogenic feature is only one facet of the dual action of HIF. Besides being a ‘guardian’ of oxygen homeostasis, HIF is capable of inducing pro-apoptotic genes¹⁴ leading to autophagy and cell death, which can be features of hypoxic tissues. In this regard, HIF can be likened to p53, which has dual roles as a guardian of genome integrity and a promoter of apoptosis.

In this review we highlight the most recently revealed features of hypoxia signalling, and the role of *hif* as a master gene controlling nutritional stress, angiogenesis, tumour metabolism, invasion and autophagy/cell death. Finally, we discuss potential new and exciting approaches to

enforce tumour regression by exploiting the emerging basic knowledge of tumour metabolism, autophagy and cell death.

Regulating angiogenesis with VEGF-A and angiopoietin-2

Growth factors and hypoxia converge in the regulation of key angiogenic genes. The cellular expansion of tumours progressively distances cells from the vasculature, and thus from oxygen and nutrients. Tumour cells, like growing embryonic cells, send out signals that initiate the formation of new blood vessels. This adaptive process, termed angiogenesis, is a general feature of every tissue; however, it is often exacerbated in solid tumours. Thus, new tumour vessels showing structural malformations, chaotic blood flow and local regions of hypoxia might nonetheless prevail. Although many molecules and receptors have been characterized in this biological process, at least two factors seem critical for initiating vessel sprouting. These are VEGF-A^{15,16} and Ang-2 (ref. 17), which are two receptor ligands expressed and secreted in response to hypoxia. VEGF-A is expressed in most cells, and attracts and guides sprouting neovessels into oxygen-depleted regions of the tumour mass^{18,19}. Endothelial cells situated at the tip of the sprouts sense and navigate through the environment using long filopodia that are rich in VEGF receptor-2 (VEGFR-2)¹⁹. Thus, migration of the tip cells is guided by a graded distribution of VEGF-A, particularly the long spliced forms that are retained in the extracellular matrix.

Although in hypoxia the binding of HIF to the *veg*f promoter is a key determinant in its expression, two other major transcriptional controls are mediated through the Ras–ERK and PI(3)K–AKT pathways^{6,20} (Fig. 1a). VEGF-A messenger RNA is upregulated by the ERK pathway through the phosphorylation of the transcription factor Sp1 and its recruitment to the proximal region of the *veg*f promoter²¹ (Fig. 1a). This regulation is independent of hypoxic stress and reflects the intensity of growth-factor stimulation or oncogenic signals. Transcriptional activation also occurs through ERK-induced phosphorylation of HIF-1α²² and the coactivator p300, which might improve the accessibility of RNA polymerase II to the *veg*f promoter. Other levels of regulation of VEGF-A occur, including the stabilization of the mRNA through the stress-activated kinase p38 (ref. 23), and translation by means of internal ribosome entry site (IRES) sequences present in 5′ non-coding regions of VEGF-A²⁴ and HIF-1α mRNAs²⁵, which are two important attributes for translation of VEGF-A under nutrient deprivation. This is another

¹Institute of Signalling, Developmental Biology and Cancer Research, CNRS UMR-6543, University of Nice, Centre Antoine Lacassagne, 33 Avenue Valombrose, 06189 Nice, France.

point of convergence between growth factors and hypoxia signalling at the level of translation. As a 'survival' cytokine, VEGF-A is translated under conditions where the cell's general translational machinery is turned off.

The second molecule induced by hypoxia is Ang-2, a receptor ligand restricted to endothelial cells^{17,26} that allows vessel remodelling by antagonizing the related molecule Ang-1. As shown in Fig. 1b, Ang-1, through Tie-2 receptor tyrosine kinase signalling and platelet-derived growth factor-B (PDGF-B) action, induces pericyte recruitment²⁷ and

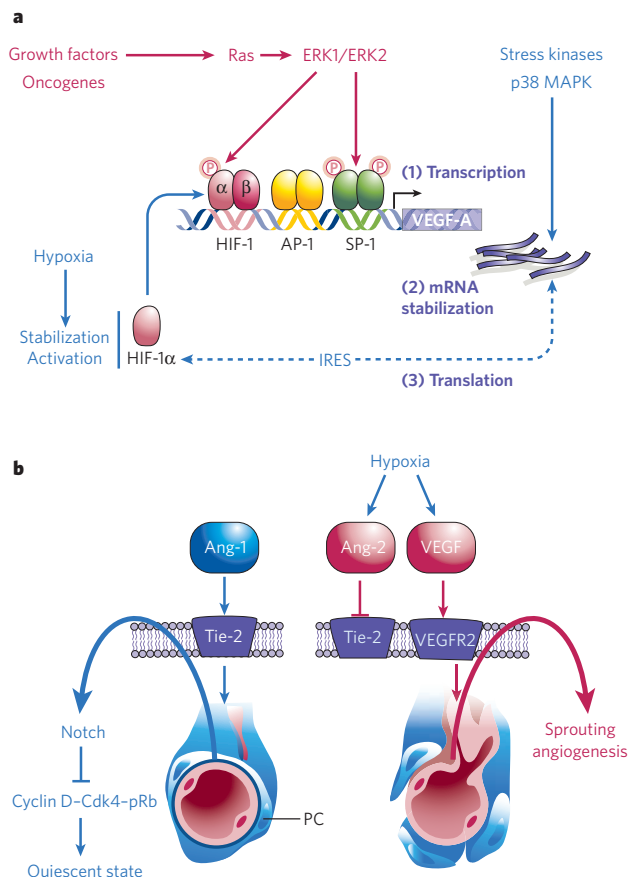


Figure 1 | VEGF-A and angiopoietin-2 are key angiogenic factors induced by hypoxia. **a**, Control of vascular endothelial growth factor-A (VEGF-A) expression. VEGF-A expression is controlled at three levels: transcription, messenger RNA stability and translation. The Ras-MEK-extracellular signal-regulated kinase (ERK) pathway stimulates transcription through phosphorylation of the transcription factors Sp1 and hypoxia-inducible factor-1α (HIF-1α) subunit, and their recruitment to the *veg* promoter. The transcription factor activator protein-1 (AP-1) might also modulate *veg* transcription. HIF-1 is a heterodimer of a hypoxia-stabilized and activated α-subunit and an oxygen-insensitive β-subunit. VEGF-A mRNA is stabilized through the stress-activated kinase p38, and the translation of VEGF-A is ensured under hypoxic and nutrient-depleted conditions by means of internal ribosome entry site (IRES) sequences. Under these energy-reduced conditions, classic cap-dependent translation is inhibited. **b**, VEGF-A and angiopoietin-2 (Ang-2) are two angiogenic factors induced by hypoxia. Blood capillaries are maintained in a mature and dormant state through the recruitment of pericytes (PC) through platelet-derived growth factor-B (PDGF-B) and the signalling of the endothelial receptor Tie-2 upon Ang-1 binding. In addition, activation of the Notch pathway through cyclin D/Cdk4 and retinoblastoma protein (pRb) phosphorylation contributes to the quiescence of endothelial cells. Ang-2 is an antagonist ligand for Tie-2 in endothelial cells and, like VEGF-A, is induced under low oxygen conditions through the HIF. The initiation of sprouting angiogenesis requires the destabilization of capillaries. This action is mediated by Ang-2, thereby blocking Tie-2 signalling and allowing VEGF-A-induced cell migration and division. MAPK, mitogen-activated protein kinase.

maturation of blood capillaries. These capillary endothelial cells are rendered quiescent through the activation of the Notch pathway²⁸, thus becoming unresponsive to VEGF-A action, unless Ang-2 is also secreted, leading to vessel destabilization. Ang-2 is a natural Ang-1 antagonist, which displaces Ang-1 from its receptor thus arresting Tie-2 signalling. Therefore Ang-2 secretion from Weibel-Palade bodies²⁹ is a critical, and perhaps limiting, step in angiogenesis permitting vessel remodeling upon VEGF-A action. It is remarkable that this angiogenic 'couple' — VEGF-A and Ang-2 — is expressed under hypoxic control when and where nutrients are needed. However, the precise mechanism of regulation of Ang-2 expression in hypoxia remains to be defined.

A master regulator of oxygen homeostasis

More than a decade ago, Semenza and colleagues showed that the nuclear factor HIF binds to the *epo* gene and induces its transcription in hypoxia¹⁰. HIF is now known to induce many genes involved in the response to hypoxia^{14,30}. HIF was shown *in vitro*, in a variety of cell-culture systems, to be activated at a cut-off point of about 5% oxygen (40 mmHg), and to progressively increase its activity with a decrease in oxygen gradient down to 0.2–0.1% oxygen (1.6–0.8 mmHg), close to anoxia. HIF belongs to the large family of basic-helix-loop-helix (bHLH) proteins and is a heterodimer of a constitutively expressed and stable HIF-1β subunit, and one of three oxygen-regulated HIF-α subunits (HIF-1α, HIF-2α or HIF-3α). HIF activation is a multi-step process involving HIF-α stabilization, nuclear translocation, heterodimerization, transcriptional activation and interaction with other proteins (see refs 31, 32 for reviews). In the nucleus, HIF binds to so-called hypoxia-response elements (HREs), with the minimal core sequence 5'-RCGTG-3', which are adjacent to auxiliary motifs specifying the responsive genes (about 100 genes have now been characterized). Yet identifying exactly how HIF becomes activated in hypoxia has been challenging.

HIF does not directly sense variations in oxygen tension (p_{O_2}). The key regulation is orchestrated by a class of 2-oxoglutarate-dependent and iron-dependent dioxygenases belonging to the largest known family of non-haem oxidizing enzymes (EC 1.14.11.2). Because the activity of these enzymes is strictly dependent on the cellular p_{O_2} , they are the true oxygen-sensing molecules controlling the hypoxic response³³. Two types of oxygen sensor control HIF action. The first are referred to as prolyl hydroxylase domain (PHD) proteins^{34–36}. PHDs hydroxylate two prolyl residues (P402 and/or P564) in the human HIF-1α region referred to as the oxygen-dependent degradation domain (ODDD; Fig. 2). This HIF-α modification specifies rapid interaction with the tumour-suppressor protein von Hippel-Lindau (VHL), a component of an E3 ubiquitin ligase complex^{37,38}. Subsequently, HIF-α subunits become marked with polyubiquitin chains that drive them to destruction by the proteasomal system^{39,40}. HIF-1α, in well-oxygenated cells (21% O_2), displays one of the shortest half-lives (<5 min) among cellular proteins⁴⁰. Of the three PHD isoforms in humans, PHD2 is the key limiting enzyme targeting HIF-1α for degradation under normoxic conditions⁴¹, whereas the physiological roles of PHD1 and PHD3, which are active under chronic hypoxia, remain to be investigated (see ref. 33 for a review).

The second type of oxygen sensor controlling the hypoxic response is an asparaginyl hydroxylase, referred to as factor inhibiting HIF-1 (FIH)⁴². This enzyme hydroxylates an asparagine residue (N803) in the most carboxy-terminal transcriptional activation domain (C-TAD) of human HIF-1α. This covalent modification abrogates C-TAD interaction with transcriptional co-activators, such as p300 and its paralogue CBP (Fig. 2). Thus, the two metabolic sensors, PHD2 and FIH, by controlling both the destruction and inactivation of HIF-α subunits, ensure full repression of the HIF pathway in well-oxygenated cells. An intriguing feature of HIF-α subunits is the occurrence of apparently 'bicephalous' transcriptional activation domains (amino (N)-TAD and C-TAD). Interestingly, only C-TAD, which interacts with p300/CBP, is subjected to hydroxylation and inhibition by FIH. We propose that N-TAD and C-TAD could discriminate between the induction of different hypoxic genes, as deletion of C-TAD in a naturally occurring spliced form of HIF-1α retains one-third of the transcriptional activity⁴³.

In a model describing the microenvironment of a blood vessel (Fig. 3), we propose that cells exposed to a decreasing oxygen gradient will first express a subset of genes regulated by N-TAD, followed by a second subset of genes regulated by C-TAD at lower oxygen concentrations. This model integrates the interesting finding that PHDs have a much lower affinity for oxygen and 2-oxoglutarate than FIH⁴⁴. Therefore, if this notion applies *in vivo*, PHDs will be inactivated at oxygen values that still maintain FIH activity and therefore keep C-TAD under repression. This model is supported by the existence of at least two classes of HIF-dependent gene: those that are sensitive and non-sensitive to the activity of FIH⁴⁵. Thus, the two TADs, together with the oxygen-sensitive discriminator FIH, constitute a cellular device that allows fine-tuning of specific HIF gene expression along the hypoxic gradient⁴⁵. One particular HIF-induced gene that seems not to be repressed by the activity of FIH is the pro-apoptotic *bnip3*, a member of the BH3-only protein family of cell death factors⁴⁶. This finding came as a surprise, because we were expecting that a gene inducing cell death should be maintained under strict and tight control, and expressed only under severe hypoxic conditions. The current hypothesis is that BNIP3-induced cell death is revealed only by severe acidosis associated with deep O₂ depletion⁴⁷ (Fig. 3). BNIP3 'activation' by acidosis requires further investigation.

HIF meets the mTOR pathway

As well as activating angiogenesis, hypoxic stress also leads to attenuation of protein synthesis by means of emerging regulatory mechanisms implicating the mTOR (mammalian target of rapamycin) pathway.

mTOR is a conserved serine/threonine protein kinase that phosphorylates a series of substrates involved in protein translation, including the eukaryotic initiation factor 4E-binding protein-1 (4EBP1) and ribosomal p70 S6 kinase (S6K)^{48,49}. As a chief orchestrator of protein synthesis, the mTOR pathway integrates a variety of signals. mTOR is activated by Rheb-GTP, a small G protein, itself negatively regulated by the GTPase activity of the tumour suppressor complex TSC2-TSC1, which was first identified as being mutated in patients with tuberous sclerosis complex (TSC)^{50,51} (Fig. 4). The upstream activators of mTOR

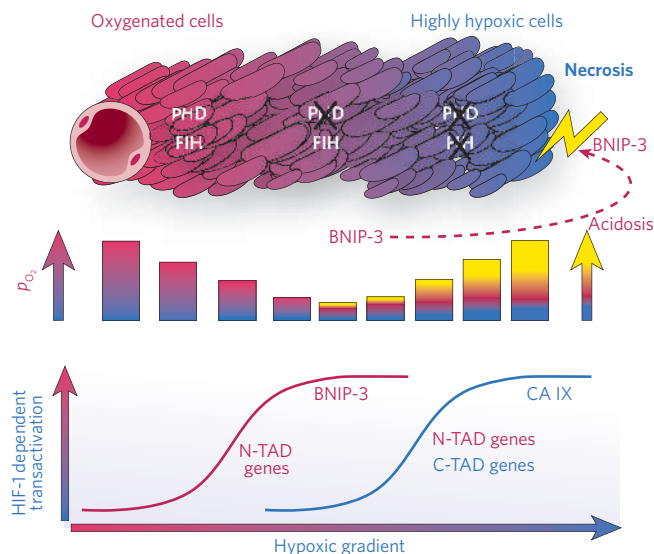


Figure 3 | Working model of two sets of HIF-1-regulated genes. In the tissue microenvironment, cells situated at various distances from blood capillaries will experience different oxygen tensions (p_{O_2}), as illustrated by a decreasing gradient. A parallel increase in the extracellular acidity due to the accumulation of lactate and CO₂ is noted as cells become more distant from capillaries. Hypoxia induces the expression of carbonic anhydrase IX (CA IX), which helps to retain a relatively neutral intracellular pH. The expression of the proapoptotic protein BNIP-3 is induced under moderately hypoxic conditions, but requires acidosis to promote cell death. Thus, under the extreme conditions of low p_{O_2} and acidosis, necrotic areas are often visible. A decreasing p_{O_2} gradient from the blood vessel to the tumour core will also determine the activity of the prolyl hydroxylase domain (PHD) proteins and factor inhibiting HIF-1 (FIH). The Michaelis constant (K_m) of the PHD proteins and FIH predict that the former has a lower affinity for oxygen and is therefore more rapidly inhibited than the latter. So, at a moderate p_{O_2} , HIF-1 α (hypoxia-inducible factor-1 α) will be stable because the PHD proteins are inhibited, but genes dependent on carboxy-terminal transcription activation domain (C-TAD) activity will not be induced because C-TAD inhibition is maintained by FIH activity. However, genes requiring only the amino-terminal transcription activation domain (N-TAD) will be induced. As the p_{O_2} decreases further, the inhibition of C-TAD will be released and HIF-1 α will attain full transcriptional activity. In this way the 'bicephalous' transcriptional nature of HIF-1 α will, in an FIH-dependent or FIH-independent manner, differentially regulate two sets of genes.

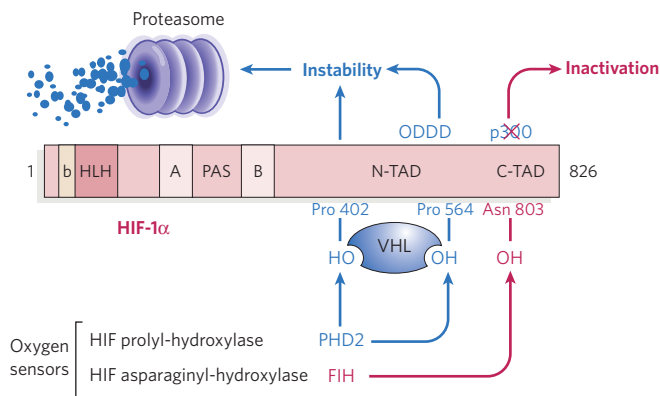


Figure 2 | Oxygen sensors contribute to the destruction and inactivation of HIF-1 α . The transcription factor HIF (hypoxia-inducible factor) is a member of the basic-helix-loop-helix PerArntSim (bHLH-PAS) family of proteins with two PAS domains, A and B. HIF is a heterodimer of an oxygen-sensitive α -subunit and an oxygen-insensitive β -subunit. Two oxygen sensors termed prolyl-hydroxylase domain (PHD) protein and factor inhibiting HIF-1 (FIH) determine, respectively, the stability and activity of HIF-1 α . The PHDs, by hydroxylating two proline residues (402 and 564) in a region called the oxygen-dependent degradation domain (ODDD), initiate the binding of a component of an E3 ubiquitin ligase, the von Hippel-Lindau (VHL) protein, which marks HIF-1 α for destruction by the proteasome. FIH, by hydroxylating an asparagine residue in the carboxy-terminal transcriptional activation domain (C-TAD) of HIF-1 α , inhibits the binding of cofactors, such as p300, that are required for the transcription of certain HIF-dependent genes. A second transcriptional activation domain, N-TAD, which overlaps the ODDD, is FIH independent and might be implicated in distinct gene expression.

are myriad growth factors, hormones and extracellular matrix components known to promote cell growth and survival through activation of the PI(3)K-AKT and Ras-ERK cardinal pathways. All these activators converge at the level of the TSC2-TSC1 integrator complex. Direct phosphorylation of TSC2, by either AKT (see the review in this issue by Shaw and Cantley, page 424) or ERKs⁵², inhibits its intrinsic GTPase activity leading to mTOR activation. Interestingly, nutrients (amino acids and glucose) also inhibit the TSC2-TSC1 complex through a mechanism that has not been fully resolved, which results in activation of protein synthesis^{53,54}. Translation of HIF-1 α has been found to be particularly sensitive to growth factors that activate mTOR^{55,56}.

In contrast to these multiple mTOR-activation inputs, mTOR, and therefore protein synthesis, is shut down under stress conditions generated by energy depletion or hypoxia (Fig. 4). In response to an increase in the AMP:ATP ratio, the upstream kinase LKB1 phosphorylates and activates AMP-activated protein kinase (AMPK)⁵⁷. Once activated, AMPK phosphorylates several downstream substrates, resulting in a decrease in energy demand by switching off ATP-consuming pathways⁵⁸. Direct phosphorylation of TSC2 by AMPK leads to activation of the TSC2-TSC1 complex and subsequent mTOR inhibition. HIF/hypoxia negatively regulates mTOR in two ways:

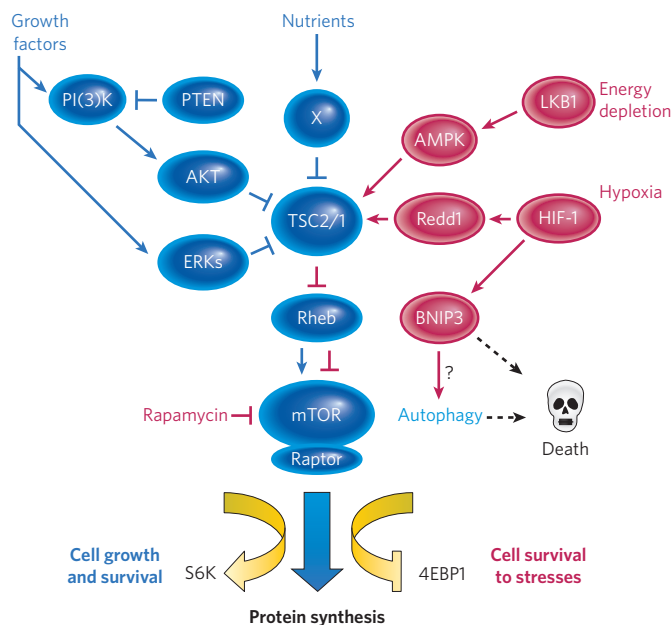


Figure 4 | Hypoxia meets the mTOR pathway. The blue arrows in the left and central parts of this diagram denote the converging pathways activating mTOR (mammalian target of rapamycin) at the level of the tumour suppressor complex (TSC2–TSC1). mTOR, which is sensitive to rapamycin, controls protein synthesis through the phosphorylation of 4E-binding protein-1 (4EBP1) and p70 S6 kinase (S6K). Growth factors, through AKT-dependent and extracellular signal-regulated kinase (ERK)-dependent phosphorylation, suppress the GTPase activity of the TSC complex, leading to full activation of GTP–Rheb, which is the activator of the mTOR–raptor complex. Nutrients and growth factors cooperate in the optimal activation of this pathway, which is essential for relaying growth and survival signals. By contrast, depletion of nutrients or energy (amino acids, ATP or oxygen) inhibits mTOR through independent activation of the TSC complex (red arrows). Suppression of mTOR in response to hypoxia requires gene induction (Redd1), whereas a decrease in ATP rapidly shuts down mTOR by activation of the AMP kinase (AMPK), thereby directly phosphorylating TSC2. The hypoxia-mediated inhibition of mTOR favours the concomitant induction of pro-apoptotic BNIP3 and macro-autophagy, which are two processes that are often associated with necrotic cell death in tumours. HIF-1, hypoxia-inducing factor-1; LKB1, serine/threonine kinase; PI(3)K, phosphatidylinositol-3-OH kinase; PTEN, phosphatase and tensin homologue.

first, hypoxia inhibits mTOR by an increase in AMP, leading to the activation of AMPK⁵⁹; and second, a more direct link between HIF and mTOR was established in *Drosophila*, in which the HIF-induced paralogue gene products, Scylla and Charybdis (REDD1/RTP801 in mammals), activate the TSC complex resulting in mTOR inhibition^{60,61}. Therefore, hypoxia, through two independent mechanisms — AMPK activation and HIF-induced REDD1 — suppresses mTOR activity (Fig. 4). These restricted nutrient conditions associated with low mTOR activity favour rapid activation of macro-autophagy, which is the ultimate survival process before cell death⁶². We believe that the rapid induction of BNIP3 in a hypoxic microenvironment contributes to cell survival via autophagy up to a point of no return, in which severe acidic areas induce necrotic cell death by means of a BNIP3-dependent action^{47,63}. Human tumour cells often silence, by promoter methylation, the expression of BNIP3 (ref. 64). Whether this BNIP3 ablation allows tumour cells to invade, metastasize and resist this nutrition-deprived and hypoxia-induced cell death requires further investigation⁶⁵.

From hypoxia to tumour invasion

Hypoxia, or genetic alterations of the hypoxia signalling cascade^{66,67} leading to the constitutive expression of HIF, could promote intense and chaotic neovascularization that facilitates tumour spread.

It has now been firmly established that HIF has important roles in tumour progression. Several immunohistochemical analyses have indicated that HIF-1 α and HIF-2 α are overexpressed in primary and metastatic human cancers, and that the level of expression, either as a result of tumour hypoxia or genetic alterations, is correlated with tumour angiogenesis and patient mortality^{14,67}. Tumour progression and invasion are often associated with the increased capacity of the cells to promote extracellular matrix remodelling, increased migration and digestion of the basement membrane. Are these key features of cancer cells regulated by HIF? From the myriad genes induced by HIF, only a limited set of gene products possesses this potential. This group includes vimentin, fibronectin, keratins 14, 18 and 19, matrix metalloproteinase 2 (MMP2), cathepsin D and urokinase plasminogen activator receptor (uPAR)¹⁴. Other factors promoting migration can be added to this list, such as the autocrine motility factor (AMF)⁶⁸, the receptor tyrosine kinase c-Met⁶⁹ and the cytokine receptor CXCR4 (ref. 70; Fig. 5).

Another landmark of invasion and a crucial feature of epithelium–mesenchyme transition (EMT) is the loss of E-cadherin expression⁷¹. Interestingly, hypoxia and genetic lesions (*vhl* deletion in renal cell carcinoma) leading to HIF activation are associated with a concomitant loss of E-cadherin. After its proposal in 1999 (ref. 72), the first evidence linking HIF to decreased expression of E-cadherin was demonstrated in ovarian carcinoma: immunolocalization of nuclear HIF-1 α showed a strong topological correlation with loss of E-cadherin⁷³. Another immunological study in early genetic lesions in kidney showed concomitant expression of carbonic anhydrase IX (CA IX), an HIF-dependent gene product, with loss of E-cadherin⁷⁴. The intriguing question is how hypoxia represses E-cadherin. E-cadherin is controlled at the transcriptional level by the labile nuclear factor family, Snail/Slug⁷¹, so that any situation leading to Snail/Slug activation will repress E-cadherin. An unexpected link has just emerged in a class of secreted enzymes, the lysyl oxidase family (LOX), which is known to modify extracellular matrix components⁷⁵ but also to have some intracellular function, such as activation of Snail. Two key lysine residues, K98 and K137, of Snail have been found to be essential in a LOXL2-induced conformational change that renders Snail partly immune to glycogen synthase kinase- β 3 (GSK- β 3)-induced degradation⁷⁶. Interestingly, high expression of LOX and LOXL2 was previously reported only in breast cancer cells with a highly invasive metastatic phenotype⁷⁷. In a separate study, stable ablation of LOX by short-hairpin RNA (shRNA) in the MDA-231 breast cancer line reduced lung and liver metastasis in mice bearing orthotopic tumours⁷⁸. Perhaps the most exciting and relevant finding for this discussion is that LOX is highly induced by HIF-1 (refs 30, 79), therefore establishing the missing link in this new molecular cascade of hypoxia signalling and metastasis: HIF–LOX–Snail–E-cadherin (Fig. 5). Besides this predicted cascade, an alternative mechanism in renal cell carcinoma has been proposed in which HIF-1 could alter the mRNA levels of the E-cadherin repressors TCF3, ZFH1A and ZFH1B⁸⁰.

Approaches to enforcing tumour regression

A strong correlation between HIF-1 α expression and increased patient mortality is well documented for many types of human cancer. In tumour–xenograft and orthotopic mouse models, manipulation of the levels of either HIF-1 α or HIF-2 α has demonstrated a causal link between HIF- α expression and tumour progression¹⁴. Perhaps the best example concerns the renal cell carcinoma line 786-0, lacking VHL, in which it was elegantly demonstrated that HIF-2 α expression was indeed causal for tumour growth⁸¹. Along this line, it is interesting to note the role of HIF-2 α in promoting cancer stem cells by specific octamer-binding transcription factor 4 (OCT-4) induction⁸². Collectively, these findings suggest that HIF, a promoter of early invasive lesions, is a promising therapeutic target in cancer. A lot of effort has been put into identifying molecules that successfully inhibit HIF-1 α expression and/or activity⁸³. However, the pleiotropic action of HIF could represent a major concern with such an approach. It might be more appropriate to identify and target the key gene products specifying HIF-dependent invasion and tumour metabolism. The first good example of targeting an HIF-

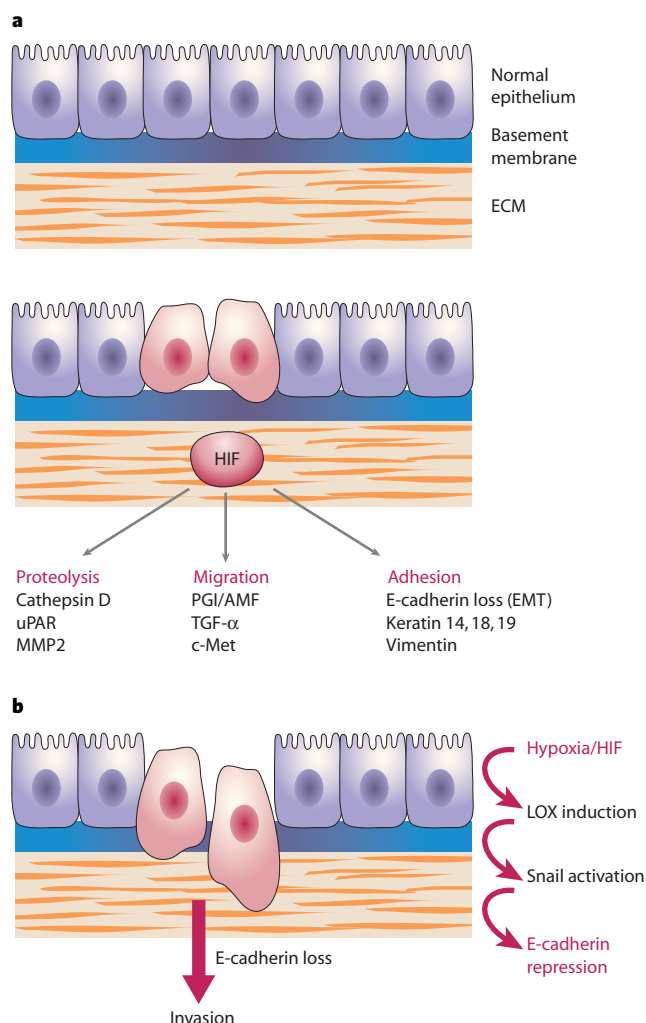


Figure 5 | Hypoxia-induced loss of E-cadherin through the lysyl oxidase-Snail activation pathway. Cell migration from the primary tumour and invasion into adjacent connective tissue are two steps leading to metastasis in carcinomas. Invasion requires a proteolytic modification of the extracellular matrix (ECM), migration and loss of cell-cell adhesion.

a, Hypoxia-inducible factor (HIF) induces markers that activate proteolysis, including cathepsin D, urokinase-type plasminogen-activator receptor (uPAR) and matrix metalloproteinase-2 (MMP2), and factors stimulating migration such as phosphoglucose isomerase/autocrine-motility factor (PGI/AMF), transforming growth factor- α (TGF- α) and the spreading factor c-Met. Epithelial-mesenchymal transition is also considered to be a trait of tumour cell invasion and is associated with a downregulation of epithelial cadherin (E-cadherin), which is essential for maintaining cell-cell adhesion. Normal epithelial cells lie on a basement membrane that makes contact with the ECM. **b**, Under low-oxygen conditions (hypoxia), HIF is active and induces the expression of lysyl oxidase (LOX). A member of the same family, LOXL2, was found to activate the transcription factor Snail, which is a strong repressor of E-cadherin. This pathway might account for the invasion and metastatic process induced by hypoxia.

dependent gene product is the anti-VEGF-A therapeutic strategy⁸⁴. Major advances have been reported recently in the treatment of metastatic colon and renal cancers with this anti-angiogenic approach, and others, such as targeting Ang-2 (ref. 85), are likely to follow.

However, every success story has its drawbacks. The concern about anti-angiogenic therapy is the inherent selection of cancer cells that adapt to more hypoxic conditions and, in particular, to tumour acidosis, which is a hallmark of anaerobic glycolysis^{86,87}. What new approaches can we propose to anticipate and counteract this drawback? In solid tumours, immunohistochemistry often shows large fronts of HIF-1 α nuclear expression delineating areas of necrosis. How can we magnify and

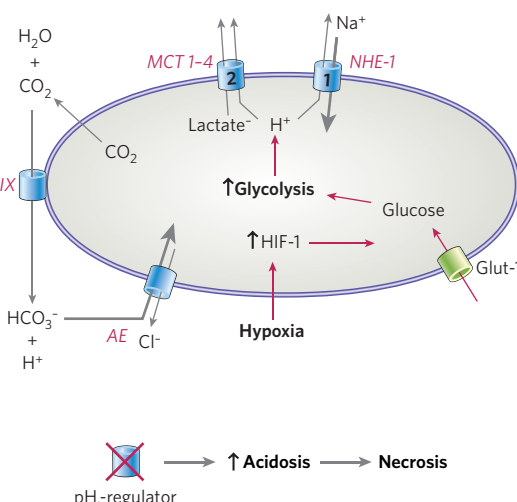


Figure 6 | Intracellular-pH-regulating systems as potential anti-cancer targets. The intracellular pH (pH_i) of tumours can become highly acidic as a result of the overproduction of lactic and carbonic acids. To survive and proliferate, cells must extrude these acids and maintain a balance between the extracellular and pH_i through the activity of several pumps, exchangers and transporters. Intracellular H⁺ ions are primarily extruded through the growth-factor-activatable and amiloride-sensitive Na⁺/H⁺ exchanger (NHE-1; target 1). Lactic acid is excreted from cells by members of the H⁺/lactate co-transporter family (monocarboxylate transporters (MCTs) 1–4; target 2). Metabolically generated CO₂ diffuses rapidly across the plasma membrane and meets the membrane-bound ectoenzyme carbonic anhydrase IX (CA IX) or CA XII, which converts it into carbonic acid (target 3). Uptake of the weak base HCO₃⁻ via a member of the Na⁺-dependent and Na⁺-independent Cl⁻/HCO₃⁻ exchangers contributes to intracellular alkalinization (target 4). Inhibiting target 1 profoundly reduces tumour incidence in the context of glycolytic tumour cells. It is expected that combined pharmacological actions on targets 1–4 will accelerate tumour necrotic cell death by intracellular acidosis. Glut-1, glucose transporter-1; HIF-1, hypoxia-inducible factor-1.

accelerate necrotic cell death in tumours? Although tumour cells have selected many mechanisms to escape and resist programmed cell death, they cannot avoid ATP depletion-induced necrosis. Several steps must be combined to enforce necrotic cell death in tumours: the first is to magnify acidosis by increased anaerobic glycolysis, the second is to antagonize pH_i regulation and the third might be to stop macro-autophagy.

The first of these steps can be conveyed through anti-angiogenic treatment, which is the best way to enforce tumour hypoxia and HIF-1 expression, and to stimulate the production of lactic acid and CO₂ in the tumour.

For the second step, we need to identify the key pH_i-regulating systems to be antagonized (representative examples are outlined in Fig. 6). Hypoxic tumour cells in particular 'learn' how to counteract local acidosis, which is a major constraint in protein synthesis and cell growth. They induce, through HIF-1 α membrane-bound ectoenzymes, CA IX or CA XII, which catalyse the hydration of CO₂ to bicarbonate⁸⁸. The rapidly formed bicarbonate is pumped in by a member of the ubiquitous bicarbonate/Cl⁻ exchanger family⁸⁹, leading to cellular alkalinization and subsequent acidification of the extracellular milieu, to the benefit of the cell. Two additional key transporters, extruding H⁺ from the cell, are the lactate/H⁺ symporters, in particular monocarboxylate transporter-4 (MCT-4) induced by HIF-1 (refs 90, 91), and the growth factor and amiloride-sensitive Na⁺/H⁺ exchanger (NHE1)^{92,93}. Preliminary experiments conducted with Ras-transformed fibroblasts deleted in *nhe1*, in glycolysis (*pgi*) or in both, have established proof of principle for tumour regression by manipulating tumour acidosis⁹⁴. Tumour cells lacking *nhe1* showed a drastic reduction in tumour incidence unless the production of lactic acid was genetically suppressed⁹⁴. An ideal

translation of this concept to tumour therapy would be to associate specific inhibitors for NHE1 (cariporide), bicarbonate/Cl⁻ exchangers (S3705)⁹⁵, MCT-1 lactate/H⁺ symporter⁹⁶ and/or blockers of carbonic anhydrases (acetazolamide).

The third step to accelerate tumour necrosis might be to suppress, in combination with other treatments, macro-autophagy⁹⁷, which is a process commonly induced by hypoxia and representing the 'ultimate' nutritional source for tumour cells to survive low-nutrient conditions⁹⁸. We expect that this 'pH_i-targeted' therapy, combined with anti-angiogenesis to increase hypoxia-mediated acidosis, will synergistically induce the collapse and massive shrinkage of solid tumours.

Preclinical studies exploiting the inducible expression of shRNA together with the above-mentioned drugs in tumour mouse models should lead to the following: the identification of the best molecular HIF-dependent targets to suppress; the recognition of the best associations of drugs targeting tumour metabolism; and the definition of the most appropriate therapeutic windows.

Concluding remarks and perspectives

The HIF signalling cascade has emerged as an important nutrient-driven genomic response allowing tumour cells to survive, expand and invade. As a result, tumour hypoxia or HIF expression is strongly associated with a diminished therapeutic response and malignant progression⁹⁹. We should therefore consider systematically assessing by immunohistochemistry both HIF-1 α /2 α and E-cadherin expression in biopsies of patients presenting with early tumour lesions. Detecting an early marker of potential malignancy and identifying patients at risk would be a step forward for clinicians, but cannot be compared with a cure. Drugs targeting HIF-1 will be available soon and there is no doubt that they will have important clinical applications^{14,83}. However, in inflammatory cells, HIF-1 serves a key function in energy metabolism and its blockade results in severe immunodeficiency¹⁰⁰. Thus, as discussed for nuclear factor- κ B (NF- κ B; see the review in this issue by Karin, page 431), prolonged and prominent inhibition of HIF-1 is unlikely to be proposed in cancer prevention. Alternatively, as emphasized in this review, it seems more appropriate to target the most relevant HIF-induced gene products/functions of the tumour microenvironment. We strongly believe that the singularities of tumour metabolism will serve as a source of inspiration for developing novel cancer therapeutic approaches.

Although the biology of hypoxia signalling has progressed rapidly, and many of the HIF-induced gene products have been characterized, their exact functions within the tumour microenvironment remain to be determined. New exciting areas in HIF signalling that intercept with the NF- κ B response and inflammation¹⁰¹, cell differentiation, Notch signalling¹⁰² and stem cell auto-renewal⁸² are blowing fresh oxygen into this area of cancer biology.

- Pages, G. *et al.* Mitogen-activated protein kinases p42^{mapk} and p44^{mapk} are required for fibroblast proliferation. *Proc. Natl Acad. Sci. USA* **90**, 8319–8323 (1993).
- Pouyssegur, J. & Lenormand, P. Fidelity and spatio-temporal control in MAP kinase (ERKs) signalling. *Eur. J. Biochem.* **270**, 3291–3299 (2003).
- Marshall, C. How do small GTPase signal transduction pathways regulate cell cycle entry? *Curr. Opin. Cell Biol.* **11**, 732–736 (1999).
- Downward, J. Mechanisms and consequences of activation of protein kinase B/Akt. *Curr. Opin. Cell Biol.* **10**, 262–267 (1998).
- Kimbara, K., Goldfinger, L. E., Hansen, M., Chou, F. L. & Ginsberg, M. H. Ras GTPases: integrins' friends or foes? *Nature Rev. Mol. Cell Biol.* **4**, 767–776 (2003).
- Rak, J. *et al.* Oncogenes and tumor angiogenesis: differential modes of vascular endothelial growth factor up-regulation in ras-transformed epithelial cells and fibroblasts. *Cancer Res.* **60**, 490–498 (2000).
- Berra, E., Pages, G. & Pouyssegur, J. MAP kinases and hypoxia in the control of VEGF expression. *Cancer Metastasis Rev.* **19**, 139–145 (2000).
- Carmeliet, P. *et al.* Role of HIF-1 α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature* **394**, 485–490 (1998).
- Pugh, C. W. & Ratcliffe, P. J. Regulation of angiogenesis by hypoxia: role of the HIF system. *Nature Med.* **9**, 677–684 (2003).
- Wang, G. L. & Semenza, G. L. Purification and characterization of hypoxia-inducible factor 1. *J. Biol. Chem.* **270**, 1230–1237 (1995).
- Semenza, G. L. Regulation of mammalian O₂ homeostasis by hypoxia-inducible factor 1. *Annu. Rev. Cell Dev. Biol.* **15**, 551–578 (1999).
- Ikeda, E., Achen, M. G., Breier, G. & Risau, W. Hypoxia-induced transcriptional activation and increased mRNA stability of vascular endothelial growth factor in C6 glioma cells. *J. Biol. Chem.* **270**, 19761–19766 (1995).
- Mandriota, S. J. & Pepper, M. S. Regulation of angiopoietin-2 mRNA levels in bovine microvascular endothelial cells by cytokines and hypoxia. *Circ. Res.* **83**, 852–859 (1998).
- Semenza, G. L. Targeting HIF-1 for cancer therapy. *Nature Rev. Cancer* **3**, 721–732 (2003).
- Ferrara, N. *et al.* Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* **380**, 439–442 (1996).
- Ferrara, N., Gerber, H. P. & LeCouter, J. The biology of VEGF and its receptors. *Nature Med.* **9**, 669–676 (2003).
- Maisonpierre, P. C. *et al.* Angiopoietin-2, a natural antagonist for Tie2 that disrupts *in vivo* angiogenesis. *Science* **277**, 55–60 (1997).
- Carmeliet, P. Blood vessels and nerves: common signals, pathways and diseases. *Nature Rev. Genet.* **4**, 710–720 (2003).
- Gerhardt, H. *et al.* VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163–1177 (2003).
- Pages, G. & Pouyssegur, J. Transcriptional regulation of the vascular endothelial growth factor gene: a concert of activating factors. *Cardiovasc. Res.* **65**, 564–573 (2005).
- Milanini-Mongiat, J., Pouyssegur, J. & Pages, G. Identification of two Sp1 phosphorylation sites for p42/p44 mitogen-activated protein kinases: their implication in vascular endothelial growth factor gene transcription. *J. Biol. Chem.* **277**, 20631–20639 (2002).
- Richard, D. E., Berra, E., Gothie, E., Roux, D. & Pouyssegur, J. p42/p44 mitogen-activated protein kinases phosphorylate hypoxia-inducible factor 1 α (HIF-1 α) and enhance the transcriptional activity of HIF-1. *J. Biol. Chem.* **274**, 32631–32637 (1999).
- Pages, G., Berra, E., Milanini, J., Levy, A. P. & Pouyssegur, J. Stress-activated protein kinases (JNK and p38/HOG) are essential for vascular endothelial growth factor mRNA stability. *J. Biol. Chem.* **275**, 26484–26491 (2000).
- Huez, I. *et al.* Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Mol. Cell Biol.* **18**, 6178–6190 (1998).
- Lang, K. J., Kappel, A. & Goodall, G. J. Hypoxia-inducible factor-1 α mRNA contains an internal ribosome entry site that allows efficient translation during normoxia and hypoxia. *Mol. Biol. Cell* **13**, 1792–1801 (2002).
- Hegen, A. *et al.* Expression of angiopoietin-2 in endothelial cells is controlled by positive and negative regulatory promoter elements. *Arterioscler. Thromb. Vasc. Biol.* **24**, 1803–1809 (2004).
- Lindblom, P. *et al.* Endothelial PDGF-B retention is required for proper investment of pericytes in the microvessel wall. *Genes Dev.* **17**, 1835–1840 (2003).
- Noseda, M. *et al.* Notch activation induces endothelial cell cycle arrest and participates in contact inhibition: role of p21Cip1 repression. *Mol. Cell Biol.* **24**, 8813–8822 (2004).
- Fiedler, U. *et al.* The Tie-2 ligand angiopoietin-2 is stored in and rapidly released upon stimulation from endothelial cell Weibel-Palade bodies. *Blood* **103**, 4150–4156 (2004).
- Manalo, D. J. *et al.* Transcriptional regulation of vascular endothelial cell responses to hypoxia by HIF-1. *Blood* **105**, 659–669 (2005).
- Brahimi-Horn, C., Mazure, N. & Pouyssegur, J. Signalling via the hypoxia-inducible factor-1 α requires multiple posttranslational modifications. *Cell Signal.* **17**, 1–9 (2005).
- Semenza, G. L. Hydroxylation of HIF-1: oxygen sensing at the molecular level. *Physiology (Bethesda)* **19**, 176–182 (2004).
- Berra, E., Ginouves, A. & Pouyssegur, J. The hypoxia-inducible-factor hydroxylases bring fresh air into hypoxia signalling. *EMBO Rep.* **7**, 41–45 (2006).
- Bruick, R. K. & McKnight, S. L. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294**, 1337–1340 (2001).
- Jaakkola, P. *et al.* Targeting of HIF- α to the von Hippel-Lindau ubiquitylation complex by O₂-regulated prolyl hydroxylation. *Science* **292**, 468–472 (2001).
- Epstein, A. C. *et al.* *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* **107**, 43–54 (2001).
- Kaelin, W. G. Jr. The von Hippel-Lindau gene, kidney cancer, and oxygen sensing. *J. Am. Soc. Nephrol.* **14**, 2703–2711 (2003).
- Maxwell, P. H., Pugh, C. W. & Ratcliffe, P. J. Activation of the HIF pathway in cancer. *Curr. Opin. Genet. Dev.* **11**, 293–299 (2001).
- Kallio, P. J., Wilson, W. J., O'Brien, S., Makino, Y. & Poellinger, L. Regulation of the hypoxia-inducible transcription factor 1 α by the ubiquitin-proteasome pathway. *J. Biol. Chem.* **274**, 6519–6525 (1999).
- Berra, E., Richard, D. E., Gothie, E. & Pouyssegur, J. HIF-1-dependent transcriptional activity is required for oxygen-mediated HIF-1 α degradation. *FEBS Lett.* **491**, 85–90 (2001).
- Berra, E. *et al.* HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1 α in normoxia. *EMBO J.* **22**, 4082–4090 (2003).
- Lando, D. *et al.* FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev.* **16**, 1466–1471 (2002).
- Gothie, E., Richard, D. E., Berra, E., Pages, G. & Pouyssegur, J. Identification of alternative spliced variants of human hypoxia-inducible factor-1 α . *J. Biol. Chem.* **275**, 6922–6927 (2000).
- Koivunen, P., Hirsila, M., Gunzler, V., Kivirikko, K. I. & Myllyharju, J. Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl-4-hydroxylase. *J. Biol. Chem.* **279**, 9899–9904 (2004).
- Dayan, F., Roux, D., Brahimi-Horn, C., Pouyssegur, J. & Mazure, N. The oxygen-sensor factor inhibiting HIF-1 (FIH) controls the expression of distinct genes through the bi-functional transcriptional character of HIF-1 α . *Cancer Res.* **66**, 3688–3698 (2006).
- Bruick, R. K. Expression of the gene encoding the proapoptotic Nip3 protein is induced by hypoxia. *Proc. Natl Acad. Sci. USA* **97**, 9082–9087 (2000).
- Webster, K. A., Graham, R. M. & Bishopric, N. H. BNP3 and signal-specific programmed death in the heart. *J. Mol. Cell Cardiol.* **38**, 35–45 (2005).
- Guertin, D. A. & Sabatini, D. M. An expanding role for mTOR in cancer. *Trends Mol. Med.* **11**, 353–361 (2005).
- Nobukuni, T. & Thomas, G. The mTOR/S6K signalling pathway: the role of the TSC1/2 tumour suppressor complex and the proto-oncogene Rheb. *Novartis Found. Symp.* **262**, 148–154; Discussion 154–159, 265–268 (2004).
- Brugarolas, J. & Kaelin, W. G. Jr. Dysregulation of HIF and VEGF is a unifying feature of the familial hamartoma syndromes. *Cancer Cell* **6**, 7–10 (2004).
- Tee, A. R., Manning, B. D., Roux, P. P., Cantley, L. C. & Blenis, J. Tuberous sclerosis complex gene products, Tuberin and Hamartin, control mTOR signaling by acting as a GTPase-activating protein complex toward Rheb. *Curr. Biol.* **13**, 1259–1268 (2003).

52. Ma, L., Chen, Z., Erdjument-Bromage, H., Tempst, P. & Pandolfi, P. P. Phosphorylation and functional inactivation of TSC2 by Erk implications for tuberous sclerosis and cancer pathogenesis. *Cell* **121**, 179–193 (2005).
53. Hara, K. *et al.* Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. *J. Biol. Chem.* **273**, 14484–14494 (1998).
54. Peng, T., Golub, T. R. & Sabatini, D. M. The immunosuppressant rapamycin mimics a starvation-like signal distinct from amino acid and glucose deprivation. *Mol. Cell Biol.* **22**, 5575–5584 (2002).
55. Fukuda, R. *et al.* Insulin-like growth factor 1 induces hypoxia-inducible factor 1-mediated vascular endothelial growth factor expression, which is dependent on MAP kinase and phosphatidylinositol 3-kinase signaling in colon cancer cells. *J. Biol. Chem.* **277**, 38205–38211 (2002).
56. Thomas, G. V. *et al.* Hypoxia-inducible factor determines sensitivity to inhibitors of mTOR in kidney cancer. *Nature Med.* **12**, 122–127 (2006).
57. Shaw, R. J. *et al.* The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress. *Proc. Natl Acad. Sci. USA* **101**, 3329–3335 (2004).
58. Kahn, B. B., Alquier, T., Carling, D. & Hardie, D. G. AMP-activated protein kinase: ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metab.* **1**, 15–25 (2005).
59. Hardie, D. G. New roles for the LKB1–AMPK pathway. *Curr. Opin. Cell Biol.* **17**, 167–173 (2005).
60. Reiling, J. H. & Hafen, E. The hypoxia-induced paralogs Scylla and Charybdis inhibit growth by down-regulating S6K activity upstream of TSC in *Drosophila*. *Genes Dev.* **18**, 2879–2892 (2004).
61. Brugarolas, J. *et al.* Regulation of mTOR function in response to hypoxia by REDD1 and the TSC1/TSC2 tumor suppressor complex. *Genes Dev.* **18**, 2893–2904 (2004).
62. Lum, J. J. *et al.* Growth factor regulation of autophagy and cell survival in the absence of apoptosis. *Cell* **120**, 237–248 (2005).
63. Bacon, A. L. & Harris, A. L. Hypoxia-inducible factors and hypoxic cell death in tumour physiology. *Ann. Med.* **36**, 530–539 (2004).
64. Okami, J., Simeone, D. M. & Logsdon, C. D. Silencing of the hypoxia-inducible cell death protein BNIP3 in pancreatic cancer. *Cancer Res.* **64**, 5338–5346 (2004).
65. Manka, D., Spicer, Z. & Millhorn, D. E. Bcl-2/adenovirus E1B 19 kDa interacting protein-3 knockdown enables growth of breast cancer metastases in the lung, liver, and bone. *Cancer Res.* **65**, 11689–11693 (2005).
66. Isaacs, J. S. *et al.* HIF overexpression correlates with biallelic loss of fumarate hydratase in renal cancer: novel role of fumarate in regulation of HIF stability. *Cancer Cell* **8**, 143–53 (2005).
67. Kim, W. Y. & Kaelin, W. G. Role of VHL gene mutation in human cancer. *J. Clin. Oncol.* **22**, 4991–5004 (2004).
68. Funasaka, T., Yanagawa, T., Hogan, V. & Raz, A. Regulation of phosphoglucose isomerase/autocrine motility factor expression by hypoxia. *FASEB J.* **19**, 1422–1430 (2005).
69. Pennacchietti, S. *et al.* Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene. *Cancer Cell* **3**, 347–361 (2003).
70. Staller, P. *et al.* Chemokine receptor CXCR4 downregulated by von Hippel–Lindau tumour suppressor pVHL. *Nature* **425**, 307–311 (2003).
71. Thiery, J. P. Epithelial–mesenchymal transitions in development and pathologies. *Curr. Opin. Cell Biol.* **15**, 740–746 (2003).
72. Beavon, I. R. Regulation of E-cadherin: does hypoxia initiate the metastatic cascade? *Mol. Pathol.* **52**, 179–188 (1999).
73. Imai, T. *et al.* Hypoxia attenuates the expression of E-cadherin via up-regulation of SNAIL in ovarian carcinoma cells. *Am. J. Pathol.* **163**, 1437–1447 (2003).
74. Esteban, M. A. *et al.* *Cancer Res.* **66**, 3567–3575 (2006).
75. Thomassin, L. *et al.* The Pro-regions of lysyl oxidase and lysyl oxidase-like 1 are required for deposition onto elastic fibers. *J. Biol. Chem.* **280**, 42848–42855 (2005).
76. Peinado, H. *et al.* A molecular role for lysyl oxidase-like 2 enzyme in snail regulation and tumor progression. *EMBO J.* **24**, 3446–3458 (2005).
77. Kirschmann, D. A. *et al.* A molecular role for lysyl oxidase in breast cancer invasion. *Cancer Res.* **62**, 4478–4483 (2002).
78. Erler, J. T. *et al.* Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* **440**, 1222–1226 (2006).
79. Denko, N. C. *et al.* Investigating hypoxic tumor physiology through gene expression patterns. *Oncogene* **22**, 5907–5914 (2003).
80. Krishnamachary, B. *et al.* Hypoxia-inducible factor-1-dependent repression of E-cadherin in von Hippel–Lindau tumor suppressor-null renal cell carcinoma mediated by TCF3, ZFX1A, and ZFX1B. *Cancer Res.* **66**, 2725–2731 (2006).
81. Kondo, K., Kico, J., Nakamura, E., Lechpammer, M. & Kaelin, W. G. Jr. Inhibition of HIF is necessary for tumor suppression by the von Hippel–Lindau protein. *Cancer Cell* **1**, 237–46 (2002).
82. Covelto, K. L. *et al.* HIF-2 α regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. *Genes Dev.* **20**, 557–570 (2006).
83. Kong, D. *et al.* Echinomycin, a small-molecule inhibitor of hypoxia-inducible factor-1 DNA-binding activity. *Cancer Res.* **65**, 9047–9055 (2005).
84. Ferrara, N. & Kerbel, R. Angiogenesis as a therapeutic target. *Nature* **438**, 967–974 (2005).
85. Oliner, J. *et al.* Suppression of angiogenesis and tumor growth by selective inhibition of angiopoietin-2. *Cancer Cell* **6**, 507–516 (2004).
86. Newell, K., Franchi, A., Pouyssegur, J. & Tannock, I. Studies with glycolysis-deficient cells suggest that production of lactic acid is not the only cause of tumor acidity. *Proc. Natl Acad. Sci. USA* **90**, 1127–1131 (1993).
87. Cardone, R. A., Casavola, V. & Reshkin, S. J. The role of disturbed pH dynamics and the Na⁺/H⁺ exchanger in metastasis. *Nature Rev. Cancer* **5**, 786–795 (2005).
88. Potter, C. & Harris, A. L. Hypoxia inducible carbonic anhydrase IX, marker of tumour hypoxia, survival pathway and therapy target. *Cell Cycle* **3**, 164–167 (2004).
89. Romero, M. F., Fulton, C. M. & Boron, W. F. The SLC4 family of HCO₃⁻-transporters. *Pflügers Arch.* **447**, 495–509 (2004).
90. Halestrap, A. P. & Meredith, D. The SLC16 gene family—from monocarboxylate transporters (MCTs) to aromatic amino acid transporters and beyond. *Pflügers Arch.* **447**, 619–628 (2004).
91. Ullah, M. S., Davies, A. J. & Halestrap, A. P. The plasma membrane lactate transporter MCT4, but not MCT1, is up-regulated by hypoxia through a HIF-1 α dependent mechanism. *J. Biol. Chem.* **281**, 9030–9037 (2006).
92. Sardet, C., Franchi, A. & Pouyssegur, J. Molecular cloning, primary structure, and expression of the human growth factor-activatable Na⁺/H⁺ antiporter. *Cell* **56**, 271–280 (1989).
93. Counillon, L. & Pouyssegur, J. The expanding family of eucaryotic Na⁺/H⁺ exchangers. *J. Biol. Chem.* **275**, 1–4 (2000).
94. Pouyssegur, J., Franchi, A. & Pages, G. pH, aerobic glycolysis and vascular endothelial growth factor in tumour growth. *Novartis Found. Symp.* **240**, 186–196; Discussion 196–198 (2001).
95. Wong, P., Kleemann, H. W. & Tannock, I. F. Cytostatic potential of novel agents that inhibit the regulation of intracellular pH. *Br. J. Cancer* **87**, 238–245 (2002).
96. Murray, C. M. *et al.* Monocarboxylate transporter MCT1 is a target for immunosuppression. *Nature Chem. Biol.* **1**, 371–376 (2005).
97. Lum, J. J., DeBerardinis, R. J. & Thompson, C. B. Autophagy in metazoans: cell survival in the land of plenty. *Nature Rev. Mol. Cell Biol.* **6**, 439–448 (2005).
98. Kondo, Y., Kanzawa, T., Sawaya, R. & Kondo, S. The role of autophagy in cancer development and response to therapy. *Nature Rev. Cancer* **5**, 726–734 (2005).
99. Vaupel, P. The role of hypoxia-induced factors in tumor progression. *Oncologist* **9** (suppl. 5), 10–17 (2004).
100. Cramer, T. *et al.* HIF-1 α is essential for myeloid cell-mediated inflammation. *Cell* **112**, 645–657 (2003).
101. Cummins, E. P. & Taylor, C. T. Hypoxia-responsive transcription factors. *Pflügers Arch.* **450**, 363–371 (2005).
102. Gustafsson, M. V. *et al.* Hypoxia requires notch signaling to maintain the undifferentiated cell state. *Dev. Cell* **9**, 617–628 (2005).

Acknowledgements We thank all our laboratory members for their discussion and support, and particularly C. Brahimi–Horn for thoroughly reviewing and critically reading the manuscript. Because of space constraints, we apologize to the many research groups whose citations were omitted or cited indirectly. Financial support was from the Centre National de la Recherche Scientifique (CNRS), Centre A. Lacassagne, Ministère de l'Éducation, de la Recherche et de la Technologie, Ligue Nationale Contre le Cancer (Equipe labellisée), the GIP HMR (contract No. 1/9743B–A3) and Conseil Régional PACA.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to J.P. (pouysseg@unice.fr).

New signals from the invasive front

Gerhard Christofori¹

Approximately 90% of all cancer deaths arise from the metastatic spread of primary tumours. Of all the processes involved in carcinogenesis, local invasion and the formation of metastases are clinically the most relevant, but they are the least well understood at the molecular level. Revealing their mechanisms is one of the main challenges for exploratory and applied cancer research. Recent experimental progress has identified a number of molecular pathways and cellular mechanisms that underlie the multistage process of metastasis formation: these include tumour invasion, tumour-cell dissemination through the bloodstream or the lymphatic system, colonization of distant organs and, finally, fatal outgrowth of metastases.

The past two decades have seen our knowledge of the genetic and epigenetic events involved in the early events of cancerogenesis increase considerably. By contrast, despite the appreciation of the clinical relevance of tumour metastasis, there is an embarrassing lack of therapies that can efficiently prevent metastasis. Multiple signal-transduction pathways, changes in the adhesive and migratory capabilities of tumour cells, and the tumour microenvironment have critical roles in malignant tumour progression. At this stage of tumour development, tumour cells migrate into and invade the surrounding tissue either as single cells or in collective clusters, thereby forming an invasive front. Here, I summarize exciting new insights into the molecular mechanisms that underpin tumour invasion and metastasis. The interactions and relationships between these processes are now being studied on various levels, leading to an increased understanding of the system involved and, it is hoped, to new cancer therapies.

Epithelial-mesenchymal transition and metastasis

About 90% of cancers originate from epithelial tissue. The most apparent morphological change that occurs during the transition from a benign tumour to a malignant and metastatic one is that tumour cells change from a highly differentiated, epithelial morphology to a migratory and invasive phenotype. Metastatic tumour cells then permeate the basal lamina barrier and invade the neighbouring tissue (Fig. 1). During this process of epithelial-mesenchymal transition (EMT), cells progressively redistribute or downregulate their apical and basolateral epithelial-specific tight and adherens junction proteins (including E-cadherin and cytokeratins) and re-express mesenchymal molecules (including vimentin and N-cadherin)¹⁻³. These changes lead to the loss of cell-cell contacts and the gain of cell motility — changes that are necessary for invasion. EMT is induced by several growth factors, which are produced either by tumour cells themselves or by stromal cells, and include transforming growth factor- β (TGF- β), hepatocyte growth factor (HGF; also known as scatter factor), epidermal growth factor (EGF), insulin-like growth factors (IGFs) and fibroblast growth factors (FGFs), and also by the upregulated proteolytic activity of matrix metalloproteases (MMPs) (Box 1). The role of EMT in the progression of epithelial cancers (carcinomas) and their metastatic dissemination is hotly debated, because in most cancers full EMT — the complete loss of epithelial markers and the gain of mesenchymal markers — is rarely observed. In fact, the presence of certain epithelial markers, such as cytokeratins, is routinely used to detect and characterize metastatic epithelial cancers in patients.

New insights into 'classical' signalling pathways

TGF- β has a dual role during tumour progression: it represses tumour growth during the early phases of tumorigenesis by inducing cell-cycle arrest and programmed cell death (apoptosis), but during the late phases of carcinogenesis, it promotes EMT, tumour invasion and metastatic dissemination of tumour cells⁴. Consistent with its tumour-suppressor functions, several components of the TGF- β -mediated signalling pathways are impaired in various human cancer types, such as TGF- β receptor II (TGF- β RII) in hereditary non-polyposis colorectal cancer, TGF- β RI in ovarian, breast and pancreatic cancers, and the signal transducers SMAD4 (also known as DPC4) and SMAD2 in pancreatic, colorectal and lung cancers. TGF- β also promotes tumour progression by exerting an immunosuppressive function. It represses the expression of major histocompatibility complex (MHC) class II and exerts negative effects on antigen-presenting cells. However, at the molecular level, the metastasis-promoting function of TGF- β is difficult to explain. Changes in components of the TGF- β signal-transduction machinery and functional synergies with other signalling pathways seem to contribute to TGF- β 's tumour-promoting function¹ (Box 1). Moreover, it induces angiogenesis by upregulating the expression of angiogenic factors, such as vascular endothelial growth factor-A (VEGF-A) and angiopoietin-1.

HGF, which is produced mainly by stromal cells, and the c-Met receptor tyrosine kinase (RTK), its cognate receptor on tumour cells, are key mediators of invasive growth during embryonic development and tumour progression⁵⁻⁷. Germline and somatic mutations, and amplifications of the *c-met* gene that lead to increased c-Met activity, are frequently found in various cancer types. Recently, c-Met activation has been shown to cause thrombohaemorrhagic events that facilitate tumour metastasis — a condition already described in the late nineteenth century as Trousseau's sign⁸.

HGF-c-Met-receptor signal transduction now serves as an example of unanticipated sophistication in RTK-mediated signalling (Box 1). In its 'classical' mode, c-Met is stimulated by its bona fide ligand, HGF, to assemble a signalling complex that induces all the signals required for cell scattering and invasive growth⁷. Besides (or perhaps as part of) its classical signal-transduction role, c-Met also selectively interacts with the cell-adhesion molecule $\alpha_6\beta_4$ integrin, which is itself phosphorylated by c-Met kinase activity (Box 1). Such phosphorylation generates additional docking sites for other adaptor molecules and signalling effectors, which, in turn, potentiate HGF-induced invasion and metastasis independently of the integrin's adhesion activity⁹. The hyaluronan receptor CD44 also cooperates with c-Met-mediated signal transduction (Box 1).

¹Department of Clinical-Biological Sciences, Center of Biomedicine, University of Basel, Mattenstrasse 28, CH-4058 Basel, Switzerland.

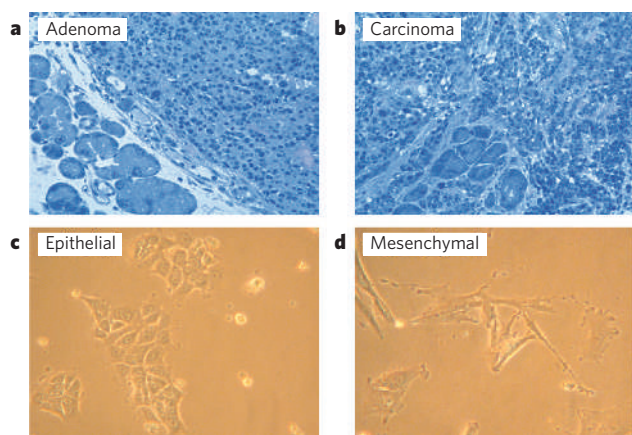


Figure 1 | The transition from epithelial tumour to invasive cancer.

a, b, Histopathology of tumours from a transgenic mouse model of pancreatic β -cell carcinogenesis (Rip1 Tag)²⁵. Note the differences between the epithelial organization of a benign adenoma (**a**) and the cell invasion and nuclear atypia of a malignant carcinoma (**b**). This transition coincides with partial epithelial–mesenchymal transition (EMT) — that is, loss of E-cadherin but not cytokeratin expression, and gain of N-cadherin but not vimentin expression (not shown). **c, d,** Cultured normal murine mammary gland (NMuMG) epithelial cells express E-cadherin and grow in epithelial-like sheets. On stimulation with TGF- β , the cells undergo full EMT — that is, they change to a mesenchymal, migratory phenotype through the loss of epithelial and the gain of mesenchymal gene expression, including the cadherin switch.

c-Met shares structural homology with plexins, which are widely expressed receptors for semaphorins (Sema). The semaphorin family consists of secreted and membrane-bound members that act mainly as guidance cues for neurons. Plexin-B1 physically associates with c-Met¹⁰. Notably, binding of Sema4D to plexin-B1 transactivates the tyrosine-kinase activity of c-Met independently of the presence of HGF¹¹. By contrast, activated Notch signalling provokes downregulation of c-Met expression and impairment of HGF-mediated signalling¹². Conversely, c-Met activation leads to the expression of Notch ligands and activation of the Notch pathway, resulting in a negative-feedback loop for fine-tuning c-Met signalling activities.

Hence, TGF- β - and HGF-induced signalling play an important part in tumour progression, notably by providing functional connections among several molecular players previously implicated in the metastatic process. They are certainly first-line targets for the development of anti-metastasis therapies. Initial proof-of-concept experiments in animal models, using soluble c-Met decoy receptors, have been highly encouraging¹³. Moreover, treatment of mice with a soluble version of TGF- β RII prevents the formation of metastases¹⁴.

New kids on the metastasis block

During the past few years, many researchers have set out to discover new genes and factors with a causal involvement in tumour metastasis. Surprisingly, many of these experiments uncovered a number of well-known genes by novel biological functions. For example, an elegant expression-cloning approach recently identified TrkB, a neurotrophic RTK for brain-derived neurotrophic factor (BDNF), as a suppressor of cell death induced by the lack of cell-matrix adhesion (anoikis)¹⁵. TrkB apparently confers survival on cells that would otherwise die when dislodged from their substrate. Instead, they migrate and invade new sites, and form metastases. TrkB and BDNF are frequently expressed in metastatic human cancers. Another example is the IGF-1 receptor (IGF-1R), which is highly expressed in many types of human cancer. Besides its important function in cell growth and survival, it promotes tumour invasion and metastasis in a transgenic mouse model of pancreatic β -cell carcinogenesis¹⁶. IGF-1R can impair E-cadherin function, thereby inducing tumour-cell migration and invasion (Box 1). IGF-1R

seems to offer an excellent multipurpose target for the development of cancer therapies. Indeed, both NVP-AEW541, a small chemical IGF-1R inhibitor, and neutralizing antibodies against IGF-1R prevent the growth of tumour cells by blocking proliferation and inducing apoptosis in culture and in animal models^{17,18}.

A much more advanced technique that is in routine clinical use is the specific repression of ErbB2 (also known as HER2) — an EGF receptor (EGFR) family member that is frequently upregulated in various cancer types, particularly breast cancer — by humanized neutralizing antibodies (Herceptin). ErbB2 expression correlates with poor prognosis and has recently been connected to tumour-cell migration by the identification of a signalling molecule, known as Momo, that relays signals to the microtubule cytoskeleton¹⁹. The hedgehog signalling pathway, another well-studied pathway that is essential in developmental patterning, has been implicated in the origin of many types of cancer. In the prostate, for example, sustained hedgehog signalling can induce tumour formation and metastasis, a process that relies on the presence of the hedgehog signal transducer smoothened. Notably, smoothened is not expressed in benign prostate epithelial cells and may, therefore, serve as a molecular marker to distinguish between benign and malignant prostate cancer²⁰.

Induction of metastatic signalling pathways by hypoxia

Tumour hypoxia not only induces tumour angiogenesis, but also modulates the expression of several genes that have been implicated in tumour metastasis. An important example is the hypoxic induction of c-met gene expression, which amplifies HGF signalling by sensitizing cells to HGF signalling²¹. Thus, hypoxia seems to affect tumour cells in two ways: it induces angiogenesis (for instance, through hypoxia-inducible factor (HIF)-1 α - and HIF-2 α -driven expression of the angiogenic factor VEGF-A) and locally adapts the tumour environment for optimal tumour growth. Moreover, by stimulating c-Met signalling and favouring tumour-cell migration and invasion, it promotes metastatic dissemination, for example, through the hypoxia- and c-Met-dependent upregulation of chemokine receptor CXCR4, which promotes breast cancer invasiveness and organ-specific metastasis^{22,23} (see below).

Changes in cell adhesion

The cadherin switch

Among the many changes in gene expression and protein function that occur during tumour progression, alterations in cell–cell and cell–matrix adhesion seem to have a central role in facilitating tumour-cell migration, invasion and metastatic dissemination. E-cadherin, the prototype member of the cadherin family of calcium-dependent cell–cell adhesion molecules, is lost concomitantly with tumour progression in most epithelial cancers²⁴, and forced downregulation of E-cadherin function in a mouse model of carcinogenesis promotes tumour invasion and metastasis²⁵.

The E-cadherin promoter is frequently repressed by specific transcriptional repressors, including Snail, Slug, SIP1, δ EF1, Twist and E12/E47, and by subsequent promoter hypermethylation^{26–29}. Some of these repressors are specifically expressed at the invasive front of human cancers, and their expression seems to be highly regulated by pathways known to promote tumour progression, including Wnt, TGF- β , FGF, EGF, signal transducer and activator of transcription 3 (STAT3) and nuclear factor- κ B (NF- κ B) signalling^{27,30,31} (Box 1). Of the transcriptional repressors, the best-studied is Snail, a highly unstable protein. It is rapidly phosphorylated by glycogen synthase kinase-3 β (GSK-3 β) and subsequently degraded by the ubiquitin–proteasome pathway³². Conversely, inhibition of GSK-3 β function results in upregulation of Snail by an NF- κ B-dependent pathway, loss of E-cadherin expression and EMT³³. Additional protein modification by lysyl oxidases 2 and 3 further stabilizes Snail protein and promotes EMT and tumour invasion³⁴.

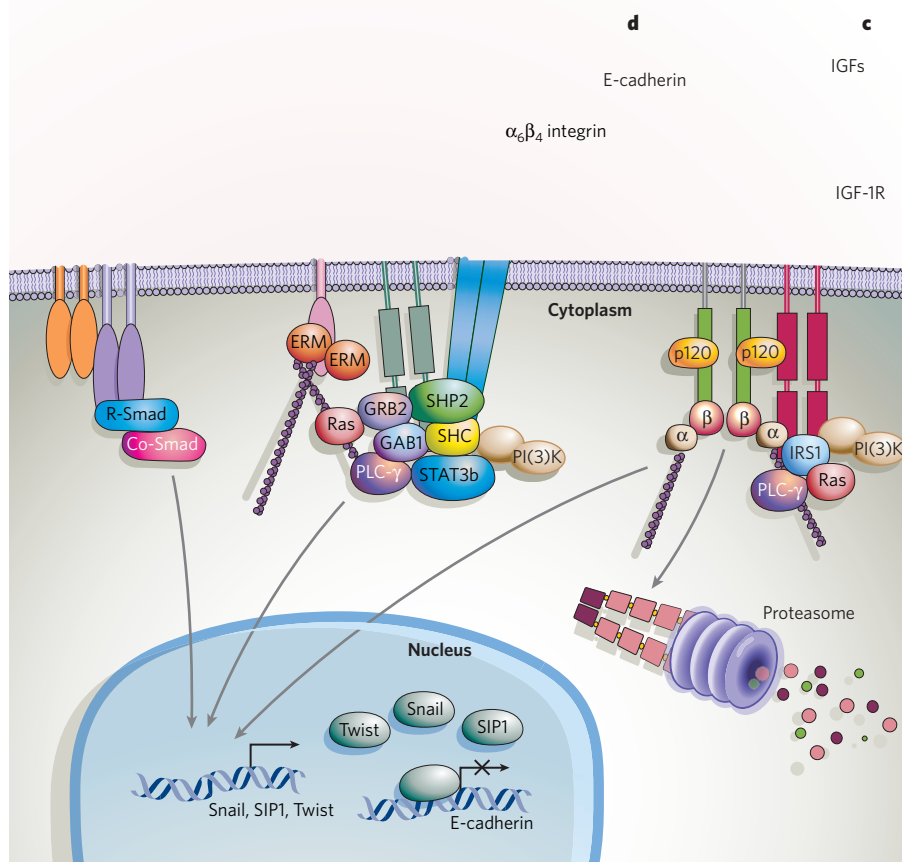
E-cadherin can also be downregulated at the protein level. RTKs, such as EGFR, c-Met, IGF-1R, FGF receptors (FGFRs) and the non-RTK c-Src can induce phosphorylation of E-cadherin and catenins, resulting in their ubiquitylation by the E3 ligase Hakai, and subsequent endocytosis and degradation³⁵ (Box 1). Finally, secreted proteases, such

Box 1 | Signalling pathways eliciting epithelial-mesenchymal transition, tumour-cell invasion and metastasis

Tumour-cell migration, invasion and metastatic dissemination all depend on changes in cell-cell and cell-matrix adhesion. E-cadherin is an important cell-adhesion molecule, and in many epithelial cancers is lost concomitantly with tumour progression. The receptor complexes shown here can all contribute to loss of E-cadherin function by means of downstream effector signalling pathways that induce the expression of transcriptional repressors of E-cadherin expression, such as Snail, SIP1 and Twist^{26,27}.

During tumour progression, TGF- β stimulates SMAD-mediated signalling by binding to and activating its receptors TGF- β RI and TGF- β RII (a). Such TGF- β - and Ras-induced late-stage tumour progression seems to depend critically on NF- κ B signalling⁷². Moreover, Par6, a regulator of epithelial-cell polarity and cell-junction assembly, interacts with TGF- β receptors and is itself a substrate for TGF- β RII kinase. Phosphorylation of Par6 is required for TGF- β -induced epithelial-mesenchymal transition, which leads to ubiquitylation of the small GTPase RhoA by the E3 ligase SMURF1, RhoA degradation by the proteasome, and a loss of tight junctions and stress fibres⁷³. Several metastasis genes are regulated by TGF- β , including those encoding CTGF, IL-11, and parathyroid-hormone-related peptide (PTHrP) — all factors that support osteoclast differentiation, osteolysis and, therefore, bone metastasis⁶⁴.

Binding of HGF to the c-Met receptor results in the assembly and activation of a signalling complex (b) consisting of the non-RTK c-Src, PI(3)K, the signalling adaptor molecules GRB2, SHC and GAB1, phospholipase C γ (PLC- γ), the tyrosine phosphatase SHP2 and the transcription factor STAT3. Binding of integrin $\alpha_6\beta_4$ to the c-Met receptor results in the integrin's phosphorylation, thereby generating additional docking sites for other adaptor molecules and signalling effectors, which potentiate HGF-induced invasion and metastasis independent of the integrin's adhesion activity⁹. c-Met also cooperates with the hyaluronan receptor CD44. Its alternatively spliced isoform CD44v6, which has previously been implicated in tumour metastasis, seems to



be required for the efficient stimulation of c-Met by HGF⁷⁴. Interestingly, the cytoplasmic tail of CD44, known to bind the actin-cytoskeleton-modulating ERM (ezrin/radixin/moesin) proteins, is essential for this function. ERM proteins link adhesion molecules to filamentous actin and participate in membrane and cytoskeletal remodelling during cell migration. CD44v3, a heparin-binding isoform of CD44, promotes HGF signalling by sequestering the HGF ligand and presenting it to c-Met receptors,

leading to increased c-Met kinase activity⁷⁵.

IGF-1R-mediated signal transduction is known to involve PI(3)K, PLC- γ , the Ras-Raf-MAPK pathway (c) and mammalian target of rapamycin (mTOR), thereby modulating protein synthesis, cell proliferation and survival (see the main text for details). IGF-1R also interacts with and phosphorylates E-cadherin and catenins, leading to their subsequent internalization and proteasomal degradation (d). α , α -catenin; β , β -catenin.

as MMPs, induced by TGF- β and HGF/SF, can cleave E-cadherin and disrupt cadherin-mediated cell-cell contacts.

What are the tumour-invasion-promoting signals elicited by the loss of E-cadherin function? First, E-cadherin loss disrupts adhesion junctions between neighbouring cells and thereby supports detachment of malignant cells from the epithelial-cell layer. Second, loss of E-cadherin has direct effects on signalling pathways involved in tumour-cell migration and tumour growth, including the canonical Wnt signalling pathway and Rho family GTPase-mediated modulation of the actin cytoskeleton (Fig. 2). However, as part of EMT, the loss of E-cadherin is frequently contrasted by the gain of expression of mesenchymal cadherins, such as N-cadherin, which enhance tumour-cell motility and migration²⁴. Hence, in addition to the loss of E-cadherin, the gain of N-cadherin (in other words, the cadherin switch) may make a critical contribution to tumour invasion and metastatic dissemination, not only by changing the adhesive repertoire of a tumour cell, but also by modulating various signalling pathways and transcriptional responses (Fig. 3).

Immunoglobulin-domain cell-adhesion molecules (IgCAMs)

Members of the immunoglobulin superfamily have critical roles in tumour progression. For example, the cell-adhesion molecule (CAM) L1 is a direct target of Wnt/ β -catenin signalling in colorectal cancer cells³⁶. It is highly expressed at the invasive front of colorectal cancers, and promotes motility, transformation and tumorigenicity in experimental cell systems. L1 is frequently co-expressed with ADAM10, a metalloprotease that cleaves L1 and causes shedding of its extracellular domain.

In a similar manner, neuronal CAM (NrCAM) is a bona fide target of Wnt signalling and confers tumorigenicity and motility on various tumour-cell types. It is highly upregulated in colorectal cancers and in melanomas³⁷. By contrast, neural CAM (NCAM) is downregulated in several types of cancer, and its loss leads to the formation of lymph node metastases in the Rip1 Tag2 transgenic mouse model of pancreatic β -cell carcinogenesis³⁸. Notably, and similarly to N-cadherin (Fig. 3), NCAM and L1 bind to and activate FGFRs in neurons and tumour cells,

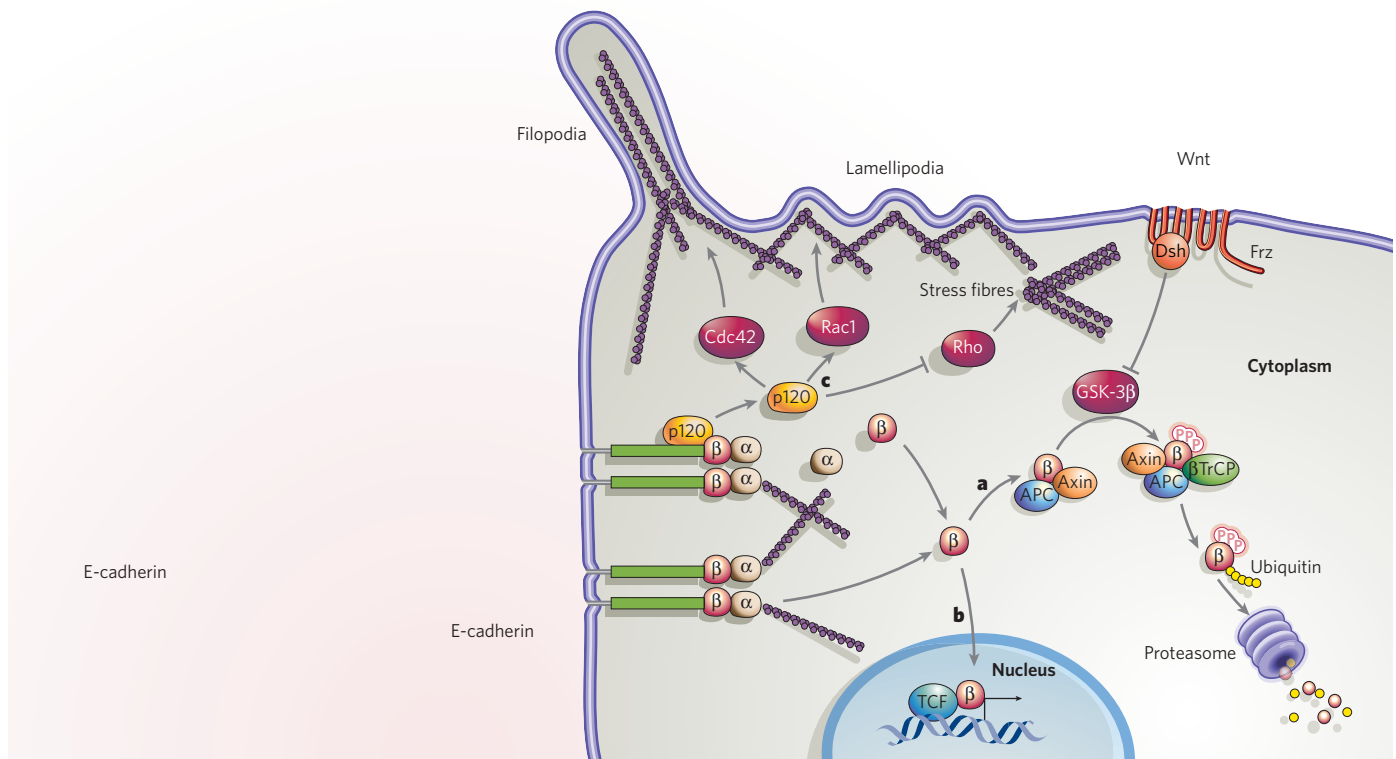


Figure 2 | Potential signalling pathways downstream of the loss of E-cadherin function. **a**, Following the loss of E-cadherin function, β -catenin (β) is sequestered by the adenomatous polyposis coli (APC)–axin–GSK-3 β complex and phosphorylated by GSK-3 β . This phosphorylated β -catenin is specifically bound and ubiquitinated by β TrCP, a subunit of the E3 ubiquitin ligase complex. Ubiquitylation earmarks β -catenin for proteasomal degradation. **b**, When the Wnt signalling pathway is activated, GSK-3 β is repressed and instead of being phosphorylated, β -catenin translocates to the nucleus. Together with TCF/Lef-1 transcription factors, it modulates the expression of several target genes involved in cell proliferation and tumour progression. **c**, On disassembly of the E-cadherin adhesion complex, displaced p120^{cas} represses the small G protein RhoA and activates Rac1 and Cdc42, which together modulate the actin cytoskeleton and the migratory behaviour of tumour cells. (Filopodia are induced by Cdc42, lamellipodia are induced by Rac1, and stress fibres are induced by RhoA.) α , α -catenin; Dsh, dishevelled; Frz, frizzled.

thereby modulating β_1 integrin-mediated cell–matrix adhesion, neurite outgrowth and cell migration³⁹.

Role of integrins

Integrins are heterodimeric cell-surface receptors that consist of two transmembrane subunits, α and β , which form distinct integrin subtypes linking extracellular matrix (ECM) ligands, such as fibronectin, vitronectin, laminin and collagen, to the intracellular actin cytoskeleton^{40,41}. Importantly, binding to these ECM components activates integrins, which, in turn, induce intracellular signalling cascades that modulate cell proliferation, survival, polarity, motility and differentiation⁴². Because there are many integrin family members with different functions, the role of integrins in tumour progression remains elusive. To detach and migrate, tumour cells depend on changes not only in cell–cell, but also in cell–matrix, interaction. Therefore, it is tacitly assumed that strong cell–matrix adhesions need to be resolved, whereas transient and weak adhesions are a prerequisite for migration. This complex situation has made it difficult to determine the functional contribution of cell–matrix adhesion to malignant tumour progression experimentally⁴². One of these pathways involving integrin functions is discussed in the context of c-Met signalling⁹ (Box 1). Another example of integrin-mediated pro-metastatic signalling was provided by the discovery of periostin, a protein that is highly expressed in metastatic colorectal cancer⁴³. Periostin seems to prevent apoptosis in both cancer cells and endothelial cells by activating the protein kinase B/Akt-mediated survival pathway through integrin $\alpha_v\beta_3$.

Contribution of the tumour microenvironment

During the past few years, results from the cellular and molecular dissection of tumour progression have led to the idea that, besides the

cellular processes and molecular pathways that exist in tumour cells themselves, an equally important contribution to malignant tumour progression comes from cells and components of the tumour micro-environment. These include endothelial cells and mural cells of blood or lymphatic vessels, tumour fibroblasts, infiltrating cells of the immune system and the tumour's ECM⁴⁴.

Carcinoma-associated fibroblasts

A key experiment for establishing the contribution of stromal fibroblasts to tumour progression was the demonstration that fibroblasts of malignant cancers are not identical to fibroblasts of the corresponding normal organ. Non-tumorigenic prostate epithelial cells co-implanted with carcinoma-associated fibroblasts (CAFs) into immunocompromised mice form tumours, whereas normal fibroblasts can impair the growth of malignant tumour cells⁴⁵. Similarly, CAFs isolated from breast carcinomas promote the growth of breast cancer cells⁴⁶. CAFs resemble myofibroblasts and affect cancer cells, at least in part, through the production of stromal-cell-derived factor 1 (SDF1), which binds its cognate receptor CXCR4 on tumour cells. CAFs are also able to stimulate tumour angiogenesis by attracting endothelial precursor cells (EPCs). Finally, expression of a dominant-negative form of TGF- β RII in stromal fibroblasts results in tumours of the prostate and forestomach, probably through the upregulated expression of HGF⁴⁷.

Cells of the immune system

The immune system has a dual function in modulating tumour progression: repression of tumour growth by the adaptive immune system (immunosurveillance) and support of tumour progression by the innate immune system (inflammatory response). But this view has been challenged. For example, the presence of T and B cells is usually taken as a

sign of active immunosurveillance of the tumour and, therefore, good prognosis. However, a recent report using a transgenic mouse model of skin carcinogenesis demonstrates that T and B cells are required for tumour malignancy⁴⁸.

The innate immune system also has a dual role in tumour progression. The presence of cells of the myeloid lineage, such as macrophages, granulocytes, neutrophils and mast cells, is indicative of good prognosis in some cancers, but seems to signify bad news in many other cancer types. In particular, monocytes and their differentiated/activated derivatives, infiltrating tumour-associated macrophages (TAMs), are found in high numbers in tumours^{49,50}. On the one hand, they secrete cytokines that support the immune response, and they promote antigen processing and presentation, resulting in an anti-tumour immune response. On the other hand, there have been many reports that the balance of secreted factors is tilted towards a cytokine/chemokine milieu that promotes tumour progression — for example, by angiogenic factors, inflammatory cytokines and MMPs⁵¹.

Bone-marrow-derived precursor cells, as well as monocytes and TAMs, express a number of receptors for ligands that are secreted by tumour cells, including VEGF receptor 1 (VEGFR1) for binding of VEGF-A and placental growth factor (PlGF), which, in turn, results in the recruitment of monocytes and TAMs to the tumour environment. The functional contribution of TAMs to tumour progression has been demonstrated — for instance, in a mouse model of skin carcinogenesis, in which tumour angiogenesis is induced by the expression of MMP9 by TAMs⁵². Conversely, depletion of macrophages represses late-stage tumour progression and metastasis, but not the development of primary tumours^{53,54}. Identification of many TAM subtypes, with potentially different functions, has just begun. For example, in experimental models, infiltrating Tie2-receptor-expressing monocytes (TEMs) promote angiogenesis and contribute to gliomagenesis⁵⁵.

Infiltrating cells of the immune system, together with CAFs and tumour cells themselves, seem to provide many pro-invasive factors, such as proteases, survival factors and angiogenic factors. Surprising insights into the potent role of proteases in tumour progression have been obtained through the transgenic expression of MMPs in breast epithelial cells, which, unexpectedly, resulted in full-blown cancer. For example, MMP3 expression results in transcriptional upregulation of the small GTPase Rac1b and upregulated levels of reactive oxygen species (ROS), which, in turn, induce expression of Snail, loss of E-cadherin expression, genomic instability and tumour progression⁵⁶. MMPs can also promote tumour-cell invasion and metastatic dissemination by activating the protease-activated receptor 1 (PAR1), a G-protein-coupled receptor implicated in metastasis of various cancer types. Although thrombin is usually the main ligand for PARs, in a xenograft mouse model of breast cancer it is predominantly MMP1, secreted by fibroblasts, that binds and activates PAR1 in cancer cells, resulting in tumour-cell migration and invasion⁵⁷.

Metastatic dissemination

Lymphogenic versus haematogenic metastasis

The formation of new blood vessels during tumour progression is a prerequisite for tumour outgrowth and can be viewed as contributing to malignant tumour progression on the basis of two main observations^{58,59}. First, angiogenesis is frequently induced by transforming signals that promote tumour progression and directly upregulate the expression of angiogenic factors. For example, VEGF-A expression is induced by the Ras–Raf–MAPK (mitogen-activated protein kinase) pathway, or by hypoxia, which also induces the expression of other proto-oncogenes, such as c-Met (see above). Second, ongoing angiogenesis and the subsequent increase in microvessel density, together with the presence of inflammatory sites, facilitate invasive tumour cells to intravasate and disseminate through the bloodstream.

Conversely, recent correlation studies in cancer patients as well as functional studies in mouse models have indicated that lymphangiogenesis, the outgrowth of new lymphatic vessels, can directly promote the formation of lymph node metastases, mainly at the draining regional lymph nodes of a tumour^{60,61}. Lymphangiogenesis is induced by the lymphangiogenic

members of the VEGF family: by binding to VEGFR3 on the surface of lymphatic endothelial cells, VEGF-C and VEGF-D stimulate the formation of new lymphatic vessels. Inflammatory responses — for example, those triggered by TNF- α , interleukin-1 β (IL-1 β) and NF- κ B signalling — seem to be involved in the regulation of VEGF-C expression. VEGF-D is the product of an immediate early, Fos-regulated gene and its expression may, therefore, be regulated by various oncogenic signalling pathways⁶⁰.

In conclusion, both angiogenesis and lymphangiogenesis contribute not only to primary tumour growth but also to the metastatic dissemination of tumour cells and, together, offer attractive targets for the development of anti-metastatic therapies. Although the first angiogenesis inhibitors are in clinical use, we need to await results on the combinatorial inhibition of angiogenesis and lymphangiogenesis, and its specific effects on tumour metastasis.

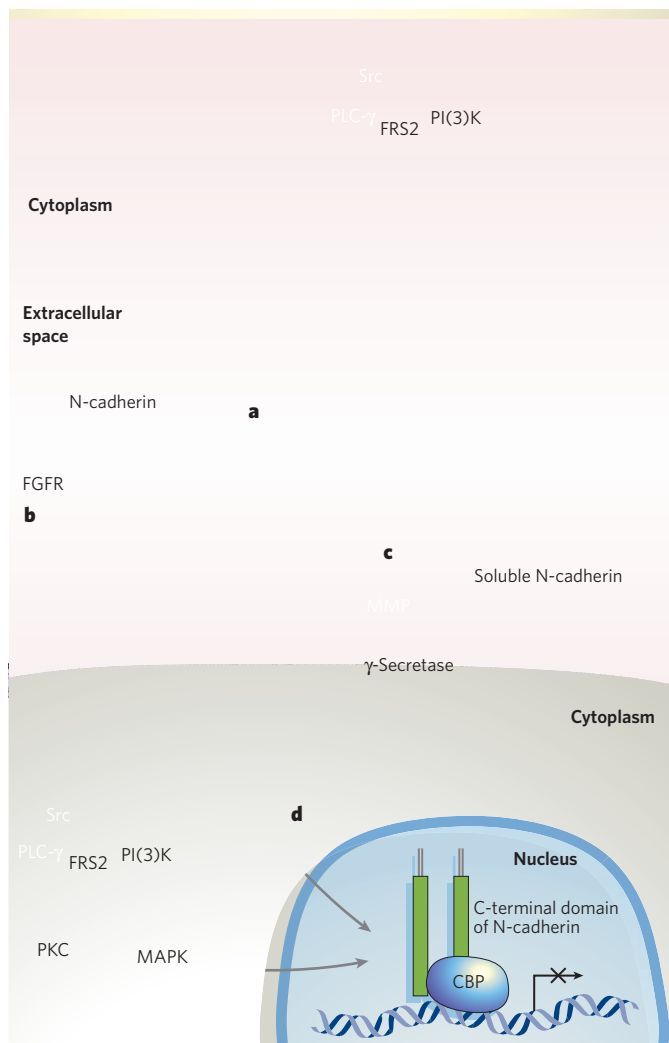


Figure 3 | The pleiotropic functions of N-cadherin. In response to loss of E-cadherin, N-cadherin is frequently upregulated during EMT (a process known as the cadherin switch). N-cadherin has several functions, all of which may contribute to tumour invasion and metastasis. **a**, Cell–cell adhesion to N-cadherin-expressing cells of the stroma. **b**, Binding and activation of FGFRs, which results in the assembly of a classical FGFR signalling complex and activation of downstream phospholipase C γ (PLC- γ), PI(3)K and MAPK signalling pathways, thereby promoting cell survival, migration and invasion^{39,70}. **c**, Cleavage and shedding of the extracellular domain of N-cadherin by MMPs. Shedded N-cadherin may neutralize N-cadherin-mediated cell–cell adhesion and/or stimulate FGFR signalling on neighbouring cells. **d**, Cleavage of N-cadherin by a γ -secretase-like protease results in translocation of the carboxy-terminal fragment of N-cadherin to the nucleus, where it represses CREB-binding protein (CBP)-mediated gene expression⁷¹.

Organ-specific metastasis

One unexplained phenomenon of tumour metastasis is the specificity with which certain cancer types metastasize to specific organs. Although the anatomy of the bloodstream and lymphatic drainage can explain some patterns of metastasis, other guiding cues seem to be involved. For example, the chemokine receptors CXCR4 and CCR7 are frequently expressed on metastatic breast cancer cells, and their ligands, SDF1/CXCL12 and CCL21, respectively, are expressed by lung and regional lymph nodes — frequent sites of breast cancer metastasis. Blockade of these ligand–receptor interactions has resulted in a reduction of metastasis in experimental mouse models⁶². In a similar manner, metastasis to the liver and lungs of a colorectal cancer cell line is impaired when CXCR4 function is blocked on tumour cells⁶³.

Recent gene-expression profiling experiments with breast cancer cell lines metastasizing to specific target organs have revealed a list of interesting genes and factors. Significant genes that are expressed in breast cancer cell lines colonizing bone include IL-11, connective-tissue growth factor (CTGF), CXCR4 and osteopontin. Combinatorial expression of these genes induces bone metastasis in cells that would not otherwise colonize bone⁶⁴. Notably, the expression of CTGF and IL-11 is under control of the TGF- β –SMAD signalling pathway described above. Genes that direct breast cancer cells to the lung include SPARC (secreted protein, acidic, cysteine-rich), Id1 (inhibitor of DNA binding 1), MMP1, MMP2, VCAM1 (vascular cell-adhesion molecule 1), IL-13R α (interleukin 13 receptor- α), COX2 (cyclooxygenase 2) and CXCL1 (ref. 65). Their combinatorial expression confers the ability to metastasize to the lungs, whereas combinatorial ablation of their expression impairs lung-specific metastasis.

Pathophysiological changes in the target organs also affect organ-specific metastasis. In a rat metastatic prostate cancer model, MMP7 and other proteases are upregulated at the bone–tumour interface and proteolytically activate the receptor activator for NF- κ B ligand (RANKL), which subsequently binds to its receptor RANK on osteoclasts. In turn, osteoclasts are activated to resorb bone and induce osteolysis, thereby releasing more MMP7 and RANKL. Consistent with this notion, MMP7-deficient mice show reduced osteolysis and impaired bone metastasis⁶⁶. The interaction between specific receptors on the surface of disseminating tumour cells and the target organ endothelium could also explain organ-specific metastasis. For example, phage-display experiments designed to identify breast cancer surface proteins that interact with the vasculature of the lung have revealed metadherin, a cell-surface protein highly expressed in breast cancer, as a functional player in lung-specific metastasis of breast cancer cells⁶⁷. A recent thought-provoking report demonstrates that bone-marrow-derived VEGFR1-positive cells play an intriguing part in directing metastatic tumour cells to a specific target organ: metastatic tumour cells secrete thus far unknown factors that induce expression of fibronectin specifically in the metastatic target organs. In turn, VEGFR1 cells are recruited to these sites, preparing the metastatic niche for the arrival of tumour cells⁶⁸. Finally, the outgrowth of secondary tumours may be another level for the design of anti-metastatic therapies, because after dissemination to specific organs, tumour cells have to adapt to and communicate with the new environment and induce angiogenesis⁶⁹.

Outlook

The many stages of development of fatal metastasis are based on a multitude of cellular and molecular mechanisms. There is good reason to believe that many of the molecular pathways and players that are causally involved may offer suitable targets for the development of efficient anti-metastatic therapies. Yet, as it stands, the identification of these players and pathways has just begun and their functional contributions remain to be explored. The ongoing development of new technologies is certainly one prerequisite for unravelling the complexity of these processes. These technologies will include high-throughput quantitative genomic and proteomic approaches combined with appropriate computational power, novel high-sensitivity and quantitative measurements of the ever-growing number of metabolic parameters and, finally, the improvement of animal models to mimic malignant human disease. ■

- Grunert, S., Jechlinger, M. & Beug, H. Diverse cellular and molecular mechanisms contribute to epithelial plasticity and metastasis. *Nature Rev. Mol. Cell Biol.* **4**, 657–665 (2003).
- Thiery, J. P. Epithelial–mesenchymal transitions in tumour progression. *Nature Rev. Cancer* **2**, 442–454 (2002).
- Huber, M. A., Kraut, N. & Beug, H. Molecular requirements for epithelial–mesenchymal transition during tumor progression. *Curr. Opin. Cell Biol.* **17**, 548–558 (2005).
- Siegel, P. M. & Massague, J. Cytostatic and apoptotic actions of TGF- β in homeostasis and cancer. *Nature Rev. Cancer* **3**, 807–821 (2003).
- Trusolino, L. & Comoglio, P. M. Scatter-factor and semaphorin receptors: cell signalling for invasive growth. *Nature Rev. Cancer* **2**, 289–300 (2002).
- Corso, S., Comoglio, P. M. & Giordano, S. Cancer therapy: can the challenge be MET? *Trends Mol. Med.* **11**, 284–292 (2005).
- Birchmeier, C., Birchmeier, W., Gherardi, E. & Vande Woude, G. F. Met, metastasis, motility and more. *Nature Rev. Mol. Cell Biol.* **4**, 915–925 (2003).
- Boccaccio, C. *et al.* The MET oncogene drives a genetic programme linking cancer to haemostasis. *Nature* **434**, 396–400 (2005).
- Trusolino, L., Bertotti, A. & Comoglio, P. M. A signaling adapter function for $\alpha 6 \beta 4$ integrin in the control of HGF-dependent invasive growth. *Cell* **107**, 643–654 (2001).
- Conrotto, P., Corso, S., Gamberini, S., Comoglio, P. M. & Giordano, S. Interplay between scatter factor receptors and B plexins controls invasive growth. *Oncogene* **23**, 5131–5137 (2004).
- Giordano, S. *et al.* The semaphorin 4D receptor controls invasive growth by coupling with Met. *Nature Cell Biol.* **4**, 720–724 (2002).
- Stella, M. C., Trusolino, L., Pennacchietti, S. & Comoglio, P. M. Negative feedback regulation of Met-dependent invasive growth by Notch. *Mol. Cell Biol.* **25**, 3982–3996 (2005).
- Michieli, P. *et al.* Targeting the tumor and its microenvironment by a dual-function decoy Met receptor. *Cancer Cell* **6**, 61–73 (2004).
- Yang, Y. A. *et al.* Lifetime exposure to a soluble TGF- β antagonist protects mice against metastasis without adverse side effects. *J. Clin. Invest.* **109**, 1607–1615 (2002).
- Douma, S. *et al.* Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB. *Nature* **430**, 1034–1039 (2004).
- Lopez, T. & Hanahan, D. Elevated levels of IGF-1 receptor convey invasive and metastatic capability in a mouse model of pancreatic islet tumorigenesis. *Cancer Cell* **1**, 339–353 (2002).
- Mitsiades, C. S. *et al.* Inhibition of the insulin-like growth factor receptor-1 tyrosine kinase activity as a therapeutic strategy for multiple myeloma, other hematologic malignancies, and solid tumors. *Cancer Cell* **5**, 221–230 (2004).
- Garcia-Echeverria, C. *et al.* In vivo antitumor activity of NVP-AEW541 — A novel, potent, and selective inhibitor of the IGF-IR kinase. *Cancer Cell* **5**, 231–239 (2004).
- Marone, R. *et al.* Memo mediates ErbB2-driven cell motility. *Nature Cell Biol.* **6**, 515–522 (2004).
- Karhadkar, S. S. *et al.* Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *Nature* **431**, 707–712 (2004).
- Pennacchietti, S. *et al.* Hypoxia promotes invasive growth by transcriptional activation of the met protooncogene. *Cancer Cell* **3**, 347–361 (2003).
- Matteucci, E., Locati, M. & Desiderio, M. A. Hepatocyte growth factor enhances CXCR4 expression favoring breast cancer cell invasiveness. *Exp. Cell Res.* **310**, 176–185 (2005).
- Phillips, R. J. *et al.* Epidermal growth factor and hypoxia-induced expression of CXCR4 chemokine receptor 4 on non-small cell lung cancer cells is regulated by the phosphatidylinositol 3-kinase/PTEN/AKT/mammalian target of rapamycin signaling pathway and activation of hypoxia inducible factor-1 α . *J. Biol. Chem.* **280**, 22473–22481 (2005).
- Cavallaro, U. & Christofori, G. Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. *Nature Rev. Cancer* **4**, 118–132 (2004).
- Perl, A. K., Wilgenbus, P., Dahl, U., Semb, H. & Christofori, G. A causal role for E-cadherin in the transition from adenoma to carcinoma. *Nature* **392**, 190–193 (1998).
- De Craene, B., van Roy, F. & Bex, G. Unraveling signalling cascades for the Snail family of transcription factors. *Cell Signal.* **17**, 535–547 (2005).
- Kang, Y. & Massague, J. Epithelial–mesenchymal transitions: twist in development and metastasis. *Cell* **118**, 277–279 (2004).
- Peinado, H., Ballestar, E., Esteller, M. & Cano, A. Snail mediates E-cadherin repression by the recruitment of the Sin3A/histone deacetylase 1 (HDAC1)/HDAC2 complex. *Mol. Cell Biol.* **24**, 306–319 (2004).
- Strathdee, G. Epigenetic versus genetic alterations in the inactivation of E-cadherin. *Semin. Cancer Biol.* **12**, 373–379 (2002).
- Yook, J. I., Li, X. Y., Ota, I., Fearon, E. R. & Weiss, S. J. Wnt-dependent regulation of the E-cadherin repressor snail. *J. Biol. Chem.* **280**, 11740–11748 (2005).
- Lu, Z., Ghosh, S., Wang, Z. & Hunter, T. Downregulation of caveolin-1 function by EGF leads to the loss of E-cadherin, increased transcriptional activity of β -catenin, and enhanced tumor cell invasion. *Cancer Cell* **4**, 499–515 (2003).
- Zhou, B. P. *et al.* Dual regulation of Snail by GSK-3 β -mediated phosphorylation in control of epithelial–mesenchymal transition. *Nature Cell Biol.* **6**, 931–940 (2004).
- Bachelder, R. E., Yoon, S. O., Franci, C., de Herreros, A. G. & Mercurio, A. M. Glycogen synthase kinase-3 is an endogenous inhibitor of Snail transcription: implications for the epithelial–mesenchymal transition. *J. Cell Biol.* **168**, 29–33 (2005).
- Peinado, H. *et al.* A molecular role for lysyl oxidase-like 2 enzyme in Snail regulation and tumor progression. *EMBO J.* **24**, 3446–3458 (2005).
- Fujita, Y. *et al.* Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nature Cell Biol.* **4**, 222–231 (2002).
- Gavert, N. *et al.* A novel target of β -catenin signaling, transforms cells and is expressed at the invasive front of colon cancers. *J. Cell Biol.* **168**, 633–642 (2005).
- Conacci-Sorell, M. E. *et al.* Nr-CAM is a target gene of the β -catenin/LEF-1 pathway in melanoma and colon cancer and its expression enhances motility and confers tumorigenesis. *Genes Dev.* **16**, 2058–2072 (2002).
- Perl, A. K. *et al.* Reduced expression of neural cell adhesion molecule induces metastatic dissemination of pancreatic β tumor cells. *Nature Med.* **5**, 286–291 (1999).
- Cavallaro, U., Niedermeyer, J., Fuxa, M. & Christofori, G. N-CAM modulates tumour-cell adhesion to matrix by inducing FGF-receptor signalling. *Nature Cell Biol.* **3**, 650–657 (2001).

40. Hood, J. D. & Cheres, D. A. Role of integrins in cell invasion and migration. *Nature Rev. Cancer* **2**, 91–100 (2002).
41. Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673–687 (2002).
42. Guo, W. & Giancotti, F. G. Integrin signalling during tumour progression. *Nature Rev. Mol. Cell Biol.* **5**, 816–826 (2004).
43. Bao, S. *et al.* Periostin potentially promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. *Cancer Cell* **5**, 329–339 (2004).
44. Bissell, M. J. & Labarge, M. A. Context, tissue plasticity, and cancer: are tumor stem cells also regulated by the microenvironment? *Cancer Cell* **7**, 17–23 (2005).
45. Hayward, S. W. *et al.* Malignant transformation in a nontumorigenic human prostatic epithelial cell line. *Cancer Res.* **61**, 8135–8142 (2001).
46. Orimo, A. *et al.* Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**, 335–348 (2005).
47. Bhowmick, N. A. *et al.* TGF- β signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science* **303**, 848–851 (2004).
48. de Visser, K. E., Korets, L. V. & Coussens, L. M. *De novo* carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. *Cancer Cell* **7**, 411–423 (2005).
49. Pollard, J. W. Tumour-educated macrophages promote tumour progression and metastasis. *Nature Rev. Cancer* **4**, 71–78 (2004).
50. Joyce, J. A. Therapeutic targeting of the tumor microenvironment. *Cancer Cell* **7**, 513–520 (2005).
51. Dranoff, G. Cytokines in cancer pathogenesis and cancer therapy. *Nature Rev. Cancer* **4**, 11–22 (2004).
52. Coussens, L. M., Tinkle, C. L., Hanahan, D. & Werb, Z. MMP-9 supplied by bone marrow-derived cells contributes to skin carcinogenesis. *Cell* **103**, 481–490 (2000).
53. Lin, E. Y., Nguyen, A. V., Russell, R. G. & Pollard, J. W. Colony-stimulating factor 1 promotes progression of mammary tumors to malignancy. *J. Exp. Med.* **193**, 727–740 (2001).
54. Aharinejad, S. *et al.* Colony-stimulating factor-1 antisense treatment suppresses growth of human tumor xenografts in mice. *Cancer Res.* **62**, 5317–5324 (2002).
55. De Palma, M. *et al.* Tie2 identifies a hematopoietic lineage of proangiogenic monocytes required for tumor vessel formation and a mesenchymal population of pericyte progenitors. *Cancer Cell* **8**, 211–226 (2005).
56. Radisky, D. C. *et al.* Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. *Nature* **436**, 123–127 (2005).
57. Boire, A. *et al.* PAR1 is a matrix metalloprotease-1 receptor that promotes invasion and tumorigenesis of breast cancer cells. *Cell* **120**, 303–313 (2005).
58. Bergers, G. & Benjamin, L. E. Tumorigenesis and the angiogenic switch. *Nature Rev. Cancer* **3**, 401–410 (2003).
59. Folkman, J. Role of angiogenesis in tumor growth and metastasis. *Semin. Oncol.* **29**, 15–18 (2002).
60. Achen, M. G., McColl, B. K. & Stacker, S. A. Focus on lymphangiogenesis in tumor metastasis. *Cancer Cell* **7**, 121–127 (2005).
61. Saharinen, P., Tammela, T., Karkkainen, M. J. & Alitalo, K. Lymphatic vasculature: development, molecular regulation and role in tumor metastasis and inflammation. *Trends Immunol.* **25**, 387–395 (2004).
62. Muller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).
63. Zeelenberg, I. S., Ruuls-Van Stalle, L. & Roos, E. The chemokine receptor CXCR4 is required for outgrowth of colon carcinoma micrometastases. *Cancer Res.* **63**, 3833–3839 (2003).
64. Kang, Y. *et al.* A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**, 537–549 (2003).
65. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).
66. Lynch, C. C. *et al.* MMP-7 promotes prostate cancer-induced osteolysis via the solubilization of RANKL. *Cancer Cell* **7**, 485–496 (2005).
67. Brown, D. M. & Ruoslahti, E. Metadherin, a cell surface protein in breast tumors that mediates lung metastasis. *Cancer Cell* **5**, 365–374 (2004).
68. Kaplan, R. N. *et al.* VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**, 820–827 (2005).
69. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nature Rev. Cancer* **2**, 563–572 (2002).
70. Suyama, K., Shapiro, I., Guttman, M. & Hazan, R. B. A signaling pathway leading to metastasis is controlled by N-cadherin and the FGF receptor. *Cancer Cell* **2**, 301–314 (2002).
71. Marambaud, P. *et al.* A CBP binding transcriptional repressor produced by the PS1/epsilon-cleavage of N-cadherin is inhibited by PS1 FAD mutations. *Cell* **114**, 635–645 (2003).
72. Huber, M. A. *et al.* NF- κ B is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. *J. Clin. Invest.* **114**, 569–581 (2004).
73. Ozdamar, B. *et al.* Regulation of the polarity protein Par6 by TGF β receptors controls epithelial cell plasticity. *Science* **307**, 1603–1609 (2005).
74. Orian-Rousseau, V., Chen, L., Sleeman, J. P., Herrlich, P. & Ponta, H. CD44 is required for two consecutive steps in HGF/c-Met signaling. *Genes Dev.* **16**, 3074–3086 (2002).
75. van der Voort, R. *et al.* Heparan sulfate-modified CD44 promotes hepatocyte growth factor/scatter factor-induced signal transduction through the receptor tyrosine kinase c-Met. *J. Biol. Chem.* **274**, 6499–6506 (1999).

Acknowledgements I apologize to all colleagues whose important work could not be cited due to space limitations. I am grateful to F. Lehenbre for providing parts of Fig. 1. Experimental work in the author's laboratory is supported by NCCR Oncology, Swiss National Science Foundation, EU-FP6 framework programme LYMPHANGIOGENOMICS LSHG-CT-2004-50357, EU-FP6 framework programme BRECOSM LSHC-CT-2004-503224, Swiss Bridge Award, Krebsliga Beider Basel, and Roche Research Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The author declares no competing financial interests. Correspondence should be addressed to G.C. (gerhard.christofori@unibas.ch).

Validating cancer drug targets

John D. Benson¹, Ying-Nan P. Chen¹, Susan A. Cornell-Kennon¹, Marion Dorsch¹, Sunkyu Kim¹, Magdalena Leszczyniecka¹, William R. Sellers¹ & Christoph Lengauer¹

A cancer drug target is only truly validated by demonstrating that a given therapeutic agent is clinically effective and acts through the target against which it was designed. Nevertheless, it is desirable to declare an early-stage drug target as 'validated' before investing in a full-scale drug discovery programme dedicated to it. Although the outcome of validation studies can guide cancer research programmes, strictly defined universal validation criteria have not been established.

Rapid advances in elucidating the molecular basis of cancer, as well as the availability of the complete sequence information of the human genome, have led to the expectation that the discovery and development of cancer drugs might become more predictable and efficient, and less serendipitous, than in the past. The key to such success is believed to lie in the concept of targeted therapy — that is, the development of drugs that influence the action or activity of a specific signalling pathway or constituent thereof. Consequently, three interrelated elements are crucial to implementing such targeted drug discovery projects. First, there must be good reason to believe that a given target (a specific gene or protein against which a drug will be developed) has causal relevance to cancer — that is, a hypothesis must be formulated with respect to the target. Second, data relevant to this hypothesis must be collected and evaluated; this includes evaluating the effects of modulating the activity of a given target in available experimental models. Third, the impact of intervention via the target must be clinically assessed. The process of evaluating potential cancer targets in this manner can be termed 'validation', and the targets that emerge might be considered, in general terms, to be 'validated'.

The manner and extent to which these strategies are integrated into cancer drug discovery and therapy are continuously evolving. This article provides a close and critical evaluation of the available tools and current rationales that might be used to tackle the challenging task of validating cancer drug targets. We initially review the conceptual framework of cancer therapy and define four tracks for the classification of cancer targets. We then scrutinize the utility of validation methods by analysing their impact on the cancer drug discovery process. We finally conclude that many different criteria might be applied when defining a validated target, and that our ability to predict the efficacy of a targeted cancer drug in the clinic holds great, but largely unfulfilled, promise.

Concepts for cancer drug therapy

The overarching hope of cancer drug discovery is to design effective and non-toxic therapies. Over time, the concept of what constitutes a cancer target has become more refined, and has been driven by both a deepening understanding of cell biology and the development of new technologies. Metabolic enzymes were the focus of drug discovery projects in the mid-twentieth century, leading to the development of folate and methotrexate as 'targeted' therapies of their day. Subsequent understanding of DNA structure and the molecular basis of DNA replication allowed therapies directed against DNA polymerases and topoisomerases to be developed. Insight into hormone signalling guided the design of cancer therapies targeting nuclear hormone receptors in breast and prostate cancer.

The deepening molecular understanding of signalling pathways has directly affected the development of targeted therapeutics. Elucidation of the roles of many kinase signalling pathways in cancer, including growth factor receptors and their effectors, along with the identification of kinases as a druggable target class, has recently been the focus of productive target-based oncology drug discovery. Moreover, discerning specific changes in cancer cells has led to more sophisticated hypotheses about the differences between them and their normal counterparts, and the relevance of these differences to the aetiology of the cancerous phenotype (Fig. 1). The quest to identify such distinguishing characteristics — whether dictated by observations of gene rearrangements or mutations, stable epigenetic changes, lineage legacies and identities, or other accrued genetic (or metabolic) liabilities — has become synonymous with modern cancer research and therapeutic development; genes for which activity, expression or dependence is thought to have increased are prime candidates for therapeutic intervention. Moreover, it is expected that future insights will be more comprehensive, accurate, sophisticated and useful than ever before. On the basis of these major principles of cancer dependencies, we can define four different subtypes of cancer target: genetics, synergy, lineage and host (Fig. 2 and Table 1).

The genetics track

The recent ability to analyse the sequence, copy number and expression levels of individual genes within cancer cells, and to simultaneously interrogate many genes in multiple independent tumours versus normal tissues, has led to the belief that such data will facilitate the robust identification of therapeutically exploitable differences between cancerous and normal cells. Such changes are believed to occur at a finite frequency in mammalian cells; serendipitous mutations confer a selective advantage to a subpopulation of cells, leading to cancer. The current conception of cancer targets assumes that tumour cells have undergone such stable changes, that at least some of the alterations are causally related to the cancer state itself (that is, they occur in oncogenes or tumour-suppressor genes), and that they are heritable between sequential cell divisions.

The first evidence of genetic alterations as causal agents in cancer was indirect, with the observation of an increased incidence of tumours in individuals or animals exposed to ionizing radiation or known mutagens. Later, the association of gross chromosomal rearrangements with leukaemias was the first direct evidence that genetic lesions might have a causal role in cancer. The first such rearrangement that was fully understood at the molecular level was a reciprocal translocation associated with chronic myeloid leukaemia (CML), which eventually became known as the Philadelphia chromosome. This translocation creates a transcript

¹Novartis Institutes for BioMedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

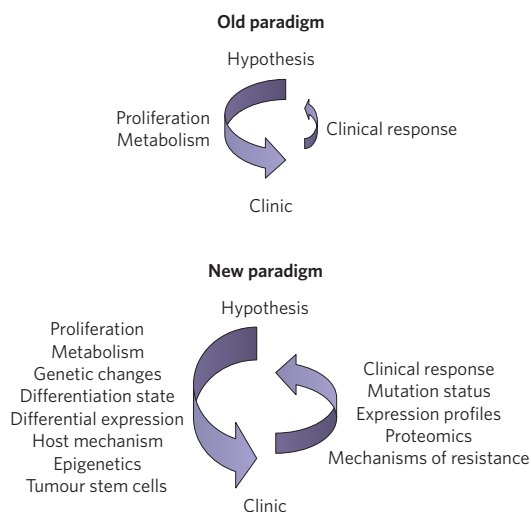


Figure 1 | Evolution of target-validation paradigms. ‘Older’ paradigms were driven by relatively empirical concepts of differences in metabolism or proliferation rates of tumour cells versus normal cells. In the absence of robust representative models, clinical response was often the sole determinant of target validation. The deepening understanding of the molecular basis of cancer, the availability of more robust experimental models and the ability to evaluate more extensively the characteristics of tumours have led to current concepts of discovery, validation and exploitation of molecular targets in cancer treatment. Importantly, such ‘new paradigms’ incorporate recent analytical technologies applicable to clinical samples, leading to greater and more insightful feedback from the clinic in evaluating available therapies. In addition, the application of such technologies now drives hypothesis formulation itself; facile identification of changes in the genetic or proteomic status of cancer cells has become a means of generating new hypotheses and of nominating promising novel therapeutic targets.

encoding a novel fusion protein known as BCR-ABL, a misregulated Abl (Abelson murine leukaemia viral oncogene) kinase in which the translocation causes replacement of the endogenous autoregulatory domain with erroneous coding sequences from BCR (breakpoint cluster region). Specific targeting of this aberrant activity led to the development of what might be considered the most notable success in the post-molecular era of cancer drug discovery: imatinib mesylate (STI571 or Gleevec). This inhibitor of the novel kinase has been highly successful in the treatment of CML, resulting in up to 80% response rates in newly treated patients¹. Whereas many translocations have subsequently been identified as genetic factors associated with leukaemias, few appear to have occurred in kinases or other proteins with enzymatic activity that would be readily addressable by small-molecule inhibitors. Moreover, no such translocation patterns have been identified that appear to be frequently associated with solid tumours.

In addition to genetic changes that alter the encoded proteins in cancer cells, stable changes in gene expression, often through the amplification of specific genes, have successfully served both as a guide for therapeutic development and as a means to identify patient cohorts that might benefit from such treatments. For example, the observation that the epidermal growth factor receptor-2 (*ERBB2*) gene is amplified, and that its encoded receptor protein is aberrantly expressed in some breast cancers triggered the development of trastuzumab (Herceptin), which is an effective antibody therapy².

Cancer cells might be dependent upon such changes for their survival; indeed, the therapeutic successes outlined above suggest that this is the case. The dependence might occur early in the evolution of a tumour, or could be contingent upon changes that occur during tumour establishment. This hypothesis of continuous dependence has become known as oncogene addiction³, and is a reasonable interpretation of the successful targeted treatments exemplified above. Methods now exist to identify with relative ease genetic changes that have occurred in cancer cells.

However, identifying the particular changes to which tumour cells have become addicted, as well as the timing of onset of such dependence, remain key challenges in applying this hypothesis to cancer therapy.

The synergy track

Synthetic lethal genetic interactions are classically defined as significantly deleterious or lethal phenotypes resulting from the combination of two or more mutations that do not produce such phenotypes individually. Gain-of-fitness changes that occur in cancer cells, allowing their survival or conferring a growth advantage, might also inadvertently sensitize the cells to other stresses that would have no consequence in normal cells but have lethal consequences in combination with the tumorigenic changes. Cells bearing such unique combinations would have distinctive liabilities that might be exploited therapeutically^{4,5}. Analogous concepts are widely applied in yeast genetics, and might also have important implications in developing specific cancer therapies.

Chemical inhibition of the function of a gene product can be considered as equivalent to a genetic loss-of-function mutation. As such, the identification of specific drug targets in cancer cells that are synthetic lethal in combination with mutant genes might allow the specific destruction of cancer cells while leaving normal cells intact. A recent example of such a scenario is the preliminary observation that inhibiting poly(ADP-ribose) polymerase (PARP) in combination with mutations that inactivate the breast cancer (*BRCA*) gene, results in tumour cell death in experimental settings⁶. Because the PARP and BRCA proteins play important roles in different DNA damage-repair processes, loss of both of these functions might have disastrous consequences for the cell. Thus, cancer cells that have undergone homozygous somatic loss of BRCA function are distinct from other cells, and might be characteristically susceptible to PARP inhibition. Rapamycin sensitivity of cells in which loss of PTEN (phosphatase and tensin homologue) or gain of phosphatidylinositol-3-OH kinase (PI(3)K) function has occurred is another example of a synthetic lethal phenotype that is applicable to cancer therapy⁷. This is consistent with the previously suggested convergence of PTEN and FRAP/mTOR (mammalian target of rapamycin) signalling in pathways involving the AKT and S6 kinases.

Along similar lines, many cancer cell lines have lost essential protective cellular mechanisms in the process of becoming tumorigenic. For example, loss of apoptotic signals, such as those mediated by B-cell leukaemia/lymphoma 2 (*BCL2*) family members, is necessary for survival of pre-malignant or tumour cells. In cases where cancer cells are tenuously poised to defy apoptotic signals, restoring the function of apoptotic signalling pathways might result in selective death of such cells. One major challenge in this area will be to determine, in a clinical setting, which tumours would be susceptible to such intervention⁸.

Although purely speculative, another aspect of synergy in cancer therapeutics relates to agents that have specific cellular targets but selectively induce the death of cancer cells by as yet unknown mechanisms. Examples include histone deacetylase (HDAC) inhibitors and heat-shock protein 90 (HSP90) inhibitors, such as geldanamycin. The observed therapeutic effects might derive from a combination of the targeted effect with other collateral changes the cancerous cell has undergone that are not present in normal cells. Although such effects are poorly defined at present, systematic exploration of such treatment modalities and their mechanisms might have great value in guiding the therapeutic use of these agents.

The lineage track

Comparative analysis of gene expression has led to the observation that patterns of gene expression from tumours of the same cell type resemble each other and their normal counterparts more closely than either cancer cells or normal cells derived from different tissue types. Thus, cancer cells often maintain many features of the cells from which they were derived. The possibility that such residual or legacy characteristics might be exploited in targeted cancer therapy has come to be known as lineage addiction. The therapeutic efficacy of oestrogen receptor antagonists, such as tamoxifen and letrozole (Femara), in the treatment

of breast cancer is a striking example of a relatively non-toxic targeted therapy that supports this concept⁹. The dependence of many prostate tumour cells on androgen receptor (AR) signalling, which is required for the survival of normal prostate secretory epithelial cells, and the therapeutic efficacy of AR antagonists are further examples of the validity of such concepts¹⁰. In addition, haematopoietic lineage has been effectively exploited in the treatment of non-Hodgkin's lymphoma: the anti-CD20 antibody rituximab (Rituxan), which recognizes a differentiation marker associated with cells of the lineage and a differentiation state characteristic of these cancer cells, has been highly effective in treating this malignancy.

In other cases, lineage status might provide a means of exploiting endogenous differentiation pathways characteristic of the tissue from which a given tumour arose, resulting in either death or re-entry of tumour cells into normal differentiation (for example, terminal differentiation). The successful use of retinoic acid in the treatment of acute promyelocytic leukaemia is an example of such a therapeutic strategy¹¹. Recent identification of the microphthalmia-associated transcription factor (*MITF*) gene as an amplified sequence in aggressive melanomas, as well as its previously established role as a differentiation and survival factor in normal melanocytes, suggests that such paradigms might be applicable to other cancers¹².

The host track

Other strategies for targeted therapy do not address tumour cells *per se*, but instead focus on tumour environment or context. The apparent requirement of many tumours to sponsor *de novo* blood vessel development has led to the recognition of angiogenesis mediators as potential therapeutic targets. Bevacizumab (Avastin), which is a monoclonal antibody that inhibits vascular endothelial growth factor (VEGF) receptor, is an example of an angiogenesis-directed therapy that has shown some efficacy in the treatment of colorectal cancers. In addition, clinical efficacy against renal cell carcinoma and gastrointestinal stromal tumours has been observed by small-molecule inhibitors such

as sorafenib (Nexavar) and sunitinib (Sutent), which target the tyrosine kinase activity of VEGF receptor¹³.

Moreover, it is possible that tumours, by evolving within a particular physiological niche, might become dependent upon certain growth factors or other environmental elements (such as stromal cells) for growth or survival. Although intuitively legitimate, such therapeutic concepts have thus far met with only limited success, perhaps because of the absence of experimental models that recapitulate the complex and perhaps idiosyncratic environments of tumours *in situ*.

Whether targets associated with metastasis represent an opportunity for therapeutic intervention is controversial and largely unexplored (see the review in this issue by Christofori, page 444). It is possible that cancer cells at a metastatic site might have dependencies or vulnerabilities distinct from the primary tumour that could be exploited therapeutically. However, as mentioned above, the establishment and validation of legitimate and representative experimental models is an important and so far unsurmounted challenge.

Validation strategies

As outlined above, the definition of a validated cancer target, and the manner in which a given target is identified and vetted, has changed over time. Our current definition of target validation is the experimental evaluation of the role of a given gene or protein in cancer; this serves as the basis for determining whether it is a promising target for cancer therapy. As such, target validation, in the strictest sense, is simply a process of hypothesis generation and testing. Criteria for evaluating the validity of a target might range from observations of altered mutation or expression status in tumours, to evidence that activity of a given target contributes to cancer cell growth in one or more experimental systems. Below, we explore in more detail three strategies for validation — a genetic approach, functional cell-based assays and validation in animal model systems (Table 2) — that can be used to assess the four types of target we have classified (Fig. 2).

Genetic validation

Patterns of somatic mutations in a specific gene in a given tumour type have come to be accepted as compelling evidence that the mutant form of the gene (and perhaps even its wild-type counterpart) has an aetiological role in that tumour type. In addition, when such mutations occur in genes associated with signalling cascades, both the mutated gene product itself and downstream effectors of that pathway are potential therapeutic targets.

One example of a signal transduction pathway in which several targets are mutated in cancer is the Ras–Raf–MEK signalling cascade. *K-RAS* is mutated in approximately 80% of pancreatic cancers. In melanoma, *N-RAS* is activated in 15% of patients, whereas *K-RAS* and *H-RAS* are rarely mutated¹⁴. The historical inability to identify effective inhibitors of Ras — either directly by targeting its GTPase function or indirectly by inhibiting farnesyl transferases that modify Ras and target it to the plasma membrane — has led to scrutiny of signalling elements downstream of *RAS* as potential targets¹⁵. A newly emerging therapeutic strategy is based upon the finding of somatic mutations of the *BRAF* gene in several different cancer types, including approximately 65% of sporadic melanomas^{16,17}. The presence of these mutations in both pre-cancerous lesions and metastatic melanoma suggests that this genetic change might be an early event in tumorigenesis¹⁸. The critical question for therapeutic development is whether tumour cells actually depend upon the activity of mutant *RAS* or *BRAF*, and, if so, at which stages of tumour development this occurs. Accordingly, validation beyond the mere presence of somatic mutations is important. Such genetic analysis of the Ras pathway has been promising: although targeting of Ras *per se* has been ineffective, inhibitors of *BRAF* and *MEK* have been developed and show early promise in the clinic¹⁹.

Epidermal growth factor receptor (EGFR) represents an interesting example of a genetically validated target. Both RNA expression levels and genetic mutations have implicated *EGFR* as a causal factor in non-small-cell lung cancer (NSCLC), leading to the development of two EGFR

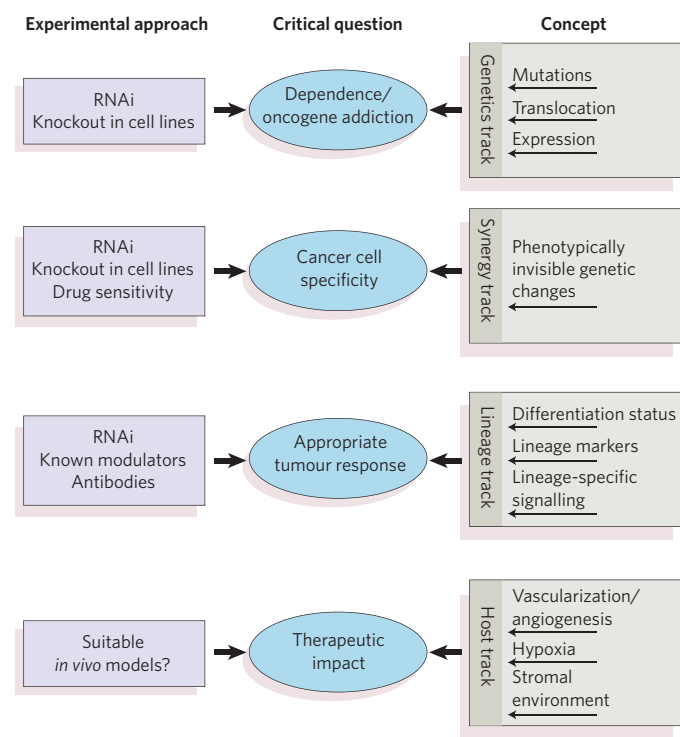


Figure 2 | Schematic overview of the four cancer drug discovery tracks. Each track is defined by the specific features listed. The critical question that must be addressed for target-validation purposes varies by track, and determines the appropriate experimental approaches for the evaluation and validation. RNAi, RNA-mediated interference.

Table 1 | Targets of approved cancer drugs

Cancer drug	Target	Disease indication	Genetic validation		Functional validation		Model validation	References
			DNA changes	Protein expression	siRNA/shRNA	Exogenous expression	Mouse models	
Genetic dependence								
Gleevec	BCR-ABL	CML, ALL	Translocation	Disregulated protein is expressed	Inhibition of cell growth and colony formation in soft agar, activation of apoptosis	BCR-ABL fusion causes transformation	Transgenic mice develop pre-B leukaemia, lymphoma	43
Gleevec	PDGFR- α , KIT	GIST	Somatic mutation, translocation	Overexpressed	Inhibition of cell proliferation	Constitutive activation of AKT and MAPK	Transplantation of KIT(G559)- and KIT(V814)-infected bone marrow cells leads to acute leukaemia	44-46
Herceptin	ERBB2	Breast cancer	Amplification, increased copy number	Overexpressed	Growth inhibition, induction of apoptosis	Constitutive activation of ERK	ERB-induced mammary tumours	47
Iressa, Tarceva	EGFR	Lung adenocarcinoma, non-small-cell lung carcinoma	Amplification, increased copy number, somatic mutation	Overexpressed	Inhibition of cell growth, induction of cell-cycle arrest, suppression of invasion		Transgenic mice develop cancer	48,49
ATRA	RAR- α	APML	Translocation	Aberrant protein is expressed		Downregulates TNF- α	NPM/RAR- α , hCG-PML/RAR- α , hCG-PLZF/RAR- α transgenic mice develop leukaemia	50,51
Lineage dependence								
Tamoxifen	Oestrogen receptor	ER+ breast cancer		Overexpressed	Reduces RAR- α , EGFR tyrosine phosphorylation and DNA synthesis	Increased proliferation	ER modulators inhibit growth and progression of pre-malignant lesions in a mouse model of ductal carcinoma <i>in situ</i>	52,53
Letrozole	Aromatase	ER+ breast cancer		Overexpressed		Increased proliferation and anchor-independent growth	Knockouts show reduced tumour incidence and delayed tumour onset	54,55
Flutamide, Bicalutamide	Androgen receptor	Prostate cancer		Overexpressed	Cell-growth inhibition	Prevents TGF- β 1-induced growth inhibition and apoptosis	Transgenic mice develop prostate cancer	56
Rituximab	CD20	Non-Hodgkin's lymphoma		B-cell marker		-	-	-
Host dependence								
Avastin	VEGF receptor	Colon cancer, pancreatic cancer		Overexpressed	Inhibition of proliferation, induction of apoptosis	-	-	57

ABL, Abelson murine leukaemia viral oncogene; AKT, protein kinase B; ALL, acute lymphoblastic leukaemia; APML, acute promyelocytic leukaemia; BCR, breakpoint cluster region; CML, chronic myeloid leukaemia; EGFR, epidermal growth factor receptor; ERK, extracellular signal-regulated kinase; ER, oestrogen receptor; GIST, gastrointestinal stromal tumours; KIT, stem cell-factor receptor; MAPK, mitogen-activated protein kinase; NPM, nucleophosmin; PDGFR- α , platelet-derived growth factor receptor- α ; PLZF, promyelocytic leukaemia zinc-finger protein; PML, promyelocytic leukaemia protein; RAR- α , retinoic acid receptor- α ; shRNA, short-hairpin RNA; siRNA, short-interfering RNA; TGF- β 1, transforming growth factor- β 1; TNF- α , tumour necrosis factor- α ; VEGF, vascular endothelial growth factor.

tyrosine kinase inhibitors: gefitinib (Iressa) and erlotinib (Tarceva). However, despite the presence of *EGFR* abnormalities in many NSCLC tumours, therapeutic inhibition of EGFR has resulted in significant tumour regression in only 10–20% of patients²⁰. Although recent studies have demonstrated the presence of activating *EGFR* mutations in responsive patients^{21,22}, more recent studies have documented response in some patients without apparent genetic alterations in *EGFR*²³. The interpretation of these clinical trials is further complicated by other factors, such as differences in drug scheduling, chemotherapy combinations and patient characteristics. Additional contributing genetic factors, such as the status of *PTEN*^{24,25} or ErbB family members²⁶, might also influence the clinical efficacy of EGFR inhibition. This serves as an important reminder of the fact that, despite the apparent validity of therapeutic hypotheses derived from genetic data, ultimate determination of the applicability of such data to clinical efficacy might occur relatively late in drug development, and could give unexpected or confounding results.

The examples cited above point out both the tremendous promise and the challenges of using genetic data in target validation. It remains difficult to predict response to therapy solely by examining expression

levels and/or the genetic status of a particular gene. However, knowing that a gene is genetically altered and selected for in cancer cells might be evidence that is compelling enough to initiate a drug discovery project. Further validation using experimental models has an important complementary role (or at least a parallel path) for understanding the function and role of genetically identified targets in cancer.

Functional target validation in cell-based systems

Experimental systems for the discovery, validation and functional analysis of tumour genes have played an important role in cancer research for several decades. The ability to introduce and express exogenous genes in non-transformed cells led to the discovery of the first transforming mutation in human cancer²⁷. The ability to express and knockout genes in mice has led to the development of model systems that have greatly advanced our understanding of the biology of cancer in a physiological setting. The more recent advent of RNA-interference (RNAi) technologies allows the 'knockdown' of expression of specific genes transiently by short-interfering RNA (siRNA), stably by short-hairpin RNA (shRNA) or in an inducible fashion via regulatable shRNA²⁸. Conceptually,

gene-specific loss of function in appropriate experimental systems should predict the outcome of using a specific inhibitory compound in tumour cells^{29,30}, and can be extended to testing of combinations of several genes or compounds in cell lines with various defined genetic backgrounds. Such approaches would ideally be used before starting a small-molecule therapeutic discovery programme, and even enable exploration of the synthetic-lethality concepts described above.

Each of these strategies has notable advantages and drawbacks. For example, in the case of knockdown-mediated target evaluation, the inability to observe any phenotype has the caveat that the threshold of lowered protein levels required to manifest a phenotype might not have been achieved. In addition, when selecting tumour cell lines that have undergone directed knockout or stable shRNA-mediated knockdown of candidate oncogenes, adaptive responses might occur (or even be selected for), leading to potentially erroneous results. Transient siRNA-mediated knockdown is neither suitable for long-term tissue-culture studies nor compatible with *in vivo*-implantation tumorigenesis models, where it might be necessary to maintain knockdown of the targeted gene for weeks or months in order to assess phenotype. In these contexts, inducible expression, homologous recombination-mediated gene knockout or shRNA systems — although challenging to implement — are particularly informative.

Cancer cell lines in tissue culture are widely utilized in early-stage evaluation of potential cancer targets, and can be used to test criteria such as growth rate, immortalization, loss of contact inhibition, two-dimensional colony formation, colony formation in soft agar and reliance upon growth factors. However, the inability to establish many tumour-derived cells in tissue culture limits the scope of these techniques. Moreover, the cells that do grow in the laboratory might have undergone adaptive changes, raising concerns as to whether they are truly representative of the cancer phenotype. Finally, conventional cell-culture systems are limited in their ability to recapitulate many aspects of the *in situ* tumour microenvironment, including hypoxia, stromal cell interaction and vascularization. Nevertheless, these systems have been valuable tools for the discovery and evaluation of potential cancer targets.

Animal model systems

Transgenic mouse models have been crucial in the study of cancers of the lymphoid and haematopoietic systems, as complex and diverse lineages of cells and cancers in these systems cannot be replicated in the laboratory. Such mouse models have served as the platform for seminal work implicating stem cells in cancer³¹.

One drawback of conventional transgenic models (that is, knockout or knock-in systems) is that the deletion of many cancer-relevant target genes leads to embryonic lethality. Moreover, although constitutive ectopic oncogene expression might lead to tumours in mice, such models do not necessarily represent initiation and progression of tumours as they arise in humans. To overcome such limitations, inducible mouse models have been developed in which the expression of certain genes can be altered at specific times.

It has become clear that many functions associated with the proliferation or survival of cancer cells are conserved in diverse organisms^{32–34}. Such functions range from cell division to regulatory signal transduction pathways. A prime example of the impact that a model organism can have on cancer research is programmed cell death (apoptosis)³⁵. Although researchers had characterized a non-necrotic form of death in cancer and other cells, the recognition of programmed cell death as a bona fide biological phenomenon was first achieved in nematodes, in which maps of cellular destiny documented the loss of specific cells at designated stages of development. Discovery of the molecular players that regulate apoptosis (including BCL2), and the finding that such players have a role in the survival of cancer cells, has dramatically changed concepts of how these cells die. Discovery of inhibitors of BCL2, as well as modulation of other apoptotic signalling and execution pathways, has become a very active area in the discovery of novel cancer therapeutics.

Future challenges

The experimental evaluation of a given targeted therapy entails consideration of both the target itself and the manner in which modulation of its function will be evaluated. This question is of overwhelming importance considering the essential roles these models play in predicting the path and outcome of targeted cancer therapy. The intrinsic value of target evaluation in model systems ultimately lies in the extent to which these systems accurately represent relevant properties of human disease, and are therefore predictive of the outcome of exploiting a given target in a therapeutic setting³⁶. For example, gene-expression profiles of biopsy samples might be expected to have more intrinsic value (that is, disease relevance) than those of cell lines derived from tumours. Whether phenotypes associated with the ability of cells to form tumours in nude mice have greater predictive value than evaluating proliferation in cell culture or colony formation in soft agar remains unclear. It is clear that current mouse models do not have sufficient predictive value³⁷, and the relative value of gene-expression data derived from patient samples versus experimental systems has not been explored systematically.

One recent example of cancer gene discovery and validation that captures the range and promise of the activities described here is the PI(3)K signalling pathway. Sequencing of the PI(3)Ks in a panel of cancer samples led to the suggestion that activating mutations in *PIK3CA* have an aetiological role in a range of human tumours^{38,39}. Indeed, Samuels and colleagues showed that deletion of activated *PI3K* alleles in colon cancer cell lines leads to reduced survival under low-serum conditions⁴⁰. Expression of such activated alleles also results in transformation of cells in tissue culture⁴¹. Thus, the finding of *PIK3CA* mutations in tumours, coupled with experimental elucidation of the activities of the encoded gene products and confirmation that mutant cell lines depend on these activities, converge to support the hypothesis that the PI(3)K pathway offers an exciting opportunity for drug discovery and therapeutic intervention in cancer.

In 2004, the US Food and Drug Administration approved only four new small molecules for treatment of cancer. Only one of these therapies came about through target-directed drug discovery; the other three therapeutics were the result of incremental improvements of agents that were originally discovered by opportunistic approaches (for example, testing compounds for effects on cell proliferation). Therefore, despite the great promise and desirability of target-directed cancer drug discovery, efforts to design effective therapeutic strategies based upon hypothesis-driven molecular targets are still in their infancy (Table 2).

The sophisticated tools now available for cancer target discovery and validation, and for subsequent development of therapeutic agents, have significantly improved the quantity and quality of information that can be collected, as well as the speed with which such data can be analysed⁴². Accordingly, the 'cycle times' for evaluating hypotheses associated with every step of cancer drug discovery and therapeutic development are becoming shorter (Fig. 1). The ability to fail or succeed more quickly suggests that progress will continue to accelerate as we learn how better to use the tools we have and to incorporate new technologies into these processes.

Table 2 | Validation strategies

Strategy	Examples
Genetic	Somatic mutation, DNA amplification RNA expression Protein expression
Cell-based	Expression of exogenous genes siRNA, shRNA knockdown Somatic knockouts/knock-ins in cancer cells Tissue culture (plastic, soft agar and so on)
Animal models	Mouse models (knockouts, inducible and transgenic) Model organisms

shRNA, short-hairpin RNA; siRNA, short-interfering RNA.

1. Deininger, M., Buchdunger, E. & Druker, B. J. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. *Blood* **1**, 2640–2653 (2005).
2. Emens, L. A. Trastuzumab: targeted therapy for the management of HER-2/neu-overexpressing metastatic breast cancer. *Am. J. Ther.* **12**, 243–253 (2005).
3. Weinstein, I. Addiction to oncogenes: the Achilles Heel of Cancer. *Science* **297**, 63–65 (2002).
4. Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Friend, S. H. Integrating genetic approaches into the discovery of anticancer drugs. *Science* **278**, 1064–1068 (1997).
5. Kaelin, W. J. Jr. The concept of synthetic lethality in the context of anticancer therapy. *Nature Rev. Cancer* **5**, 689–698 (2005).
6. Farmer, H. et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).
7. Mills, G. P., Lu, Y. & Kohn, E. C. Linking molecular therapeutics to molecular diagnostics: inhibition of the FRAP/RAFT/TOR component of the PI3K pathway preferentially blocks PTEN mutant cells *in vitro* and *in vivo*. *Proc. Natl Acad. Sci. USA* **98**, 10031–10033 (2001).
8. O'Neill, J., Manion, M., Schwartz, P. & Hockenbery, D. M. Promises and challenges of targeting Bcl-2 anti-apoptotic proteins for cancer therapy. *Biochim. Biophys. Acta* **1705**, 43–51 (2004).
9. Jensen, E. V. & Jordan, V. C. The estrogen receptor: a model for molecular medicine. *Clin. Cancer Res.* **9**, 1980–1989 (2003).
10. Jenster, G. The role of the androgen receptor in the development and progression of prostate cancer. *Semin. Oncol.* **26**, 407–421 (1999).
11. Pitha-Rowe, I., Petty, W. J., Kitareewan, S. & Dmitrovsky, E. Retinoid target genes in acute promyelocytic leukemia. *Leukemia* **17**, 1723–1730 (2003).
12. Garraway, L. A. et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
13. Marx, J. Encouraging results for second generation anti-angiogenic drugs. *Science* **308**, 1248–1249 (2005).
14. van Elsas, A. et al. Relevance of ultraviolet-induced *N-ras* oncogene point mutations in development of primary human cutaneous melanoma. *Am. J. Pathol.* **149**, 883–893 (1996).
15. Walker, F. & Olson, M. F. Targeting Ras and Rho GTPases as opportunities for cancer therapies. *Curr. Opin. Genet. Dev.* **15**, 62–68 (2005).
16. Beeram, M., Patnaik, A. & Rowinsky, E. K. Raf: a strategic target for therapeutic development against cancer. *J. Clin. Oncol.* **23**, 6771–6790 (2005).
17. Davies, H. et al. Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
18. Pollock, P. M. et al. High frequency of *BRAF* mutations in nevi. *Nature Genet.* **33**, 19–20 (2003).
19. Thompson, N. & Lyons, J. Recent progress in targeting the Raf/MEK/ERK pathway with inhibitors in cancer drug discovery. *Curr. Opin. Pharmacol.* **5**, 350–356 (2005).
20. Janne, P. A., Engelman, J. A. & Johnson, B. E. Epidermal growth factor receptor mutations in non-small-cell lung cancer: implications for treatment and tumor biology. *J. Clin. Oncol.* **23**, 3227–3234 (2005).
21. Paez, J. G. et al. *EGFR* mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–500 (2004).
22. Rao, W. et al. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl Acad. Sci. USA* **101**, 13306–13311 (2004).
23. Tsao, M. S. et al. Erlotinib in lung cancer: molecular and clinical predictors of outcome. *N. Engl. J. Med.* **353**, 133–134 (2005).
24. Patel, J. D., Pasche, B. & Argiris, A. Targeting non-small cell lung cancer with epidermal growth factor tyrosine kinase inhibitors: where do we stand, where do we go. *Crit. Rev. Oncol. Hematol.* **50**, 175–186 (2004).
25. Mellinghoff, I. K. et al. Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *N. Engl. J. Med.* **353**, 2012–2024 (2005).
26. Engelman, J. A. et al. ErbB-3 mediates phosphoinositide 3-kinase activity in gefitinib-sensitive non-small cell lung cancer cell lines. *Proc. Natl Acad. Sci. USA* **102**, 3788–3793 (2005).
27. Shih, C. & Weinberg, R. A. Isolation of a transforming sequence from a human bladder carcinoma cell line. *Cell* **29**, 161–169 (1982).
28. Silva, J., Chang, K., Hannon, G. J. & Rivas, F. V. RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age. *Oncogene* **23**, 8401–8409 (2004).
29. Torrance, C. J., Agrawal, V., Vogelstein, B. & Kinzler, K. W. Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nature Biotechnol.* **19**, 940–945 (2001).
30. Kohli, M., Rago, C., Lengauer, C., Kinzler, K. W. & Vogelstein, B. Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. *Nucleic Acids Res.* **32**, e3 (2004).
31. Mazurier, F., Doedens, M., Gan, O. I. & Dick, J. E. Rapid myeloerythroid repopulation after intrafemoral transplantation of NOD-SCID mice reveals a new class of human stem cells. *Nature Med.* **9**, 959–963 (2003).
32. Brumby, A. M. & Richardson, H. E. Using *Drosophila melanogaster* to map human cancer pathways. *Nature Rev. Cancer* **5**, 626–639 (2005).
33. Amatrua, J. D., Shepard, J. L., Stern, H. M. & Zon, L. I. Zebrafish as a cancer model system. *Cancer Cell* **1**, 229–231 (2002).
34. Simon, J. A. Yeast as a model system for anticancer drug discovery. *Expert Opin. Ther. Targets* **5**, 177–195 (2001).
35. Hengartner, M. O. & Horvitz, H. R. C. *elegans* cell survival gene *ced-9* encodes a functional homolog of the mammalian proto-oncogene *bcl-2*. *Cell* **76**, 665–676 (1994).
36. Kamb, A. What's wrong with our cancer models? *Nature Rev. Drug Discov.* **4**, 161–165 (2005).
37. Sausville, E. A. Target selection issues in drug discovery and development. *J. Chemother. Suppl.* **16** (suppl. 4), 16–18 (2004).
38. Samuels, Y. et al. High frequency of mutations of the *PIK3CA* gene in human cancers. *Science* **304**, 554 (2004).
39. Bachman, K. E. et al. The *PIK3CA* gene is mutated with high frequency in human breast cancers. *Cancer Biol. Ther.* **3**, 772–775 (2004).
40. Samuels, Y. et al. Mutant *PIK3CA* promotes cell growth and invasion of human cancer cells. *Cancer Cell* **7**, 561–573 (2005).
41. Aoki, M., Jiang, H. & Vogt, P. K. Proteasomal degradation of the FoxO1 transcriptional regulator in cells transformed by the P3k and Akt oncoproteins. *Proc. Natl Acad. Sci. USA* **101**, 13613–13617 (2004).
42. Lengauer, C., Diaz, L. A. & Saha, S. Cancer drug discovery through collaboration. *Nature Rev. Drug Discov.* **4**, 375–380 (2005).
43. Sallouk, H. F., Vowles, I., Heisterkamp, N., Groffen, J. & Laneuville, P. Early events in leukemogenesis in P190Bcr-abl transgenic mice. *Oncogene* **19**, 4362–4374 (2000).
44. Lefevre, G. et al. Roles of stem cell factor/c-Kit and effects of Glivec/STI571 in human uveal melanoma cell tumorigenesis. *J. Biol. Chem.* **279**, 31769–31779 (2004).
45. Mitsiades, C. S., Mitsiades, N. & Koutsilieris, M. The Akt pathway: molecular targets for anti-cancer drug development. *Curr. Cancer Drug Targets* **4**, 235–256 (2004).
46. Kitayama, H. et al. Neoplastic transformation of normal hematopoietic cells by constitutively activating mutations of c-kit receptor tyrosine kinase. *Blood* **88**, 995–1004 (1996).
47. Cool, M., Depault, F. & Jolicoeur, P. Fine allelotyping of ErbB2-induced mammary tumors in mice reveals multiple discontinuous candidate regions of tumor-suppressor loci. *Genes Chromosomes Cancer* **45**, 191–202 (2006).
48. Kang, C. S. et al. An *in vitro* study on the suppressive effect of glioma cell growth induced by plasmid-based small interference RNA (siRNA) targeting human epidermal growth factor receptor. *J. Neurooncol.* **73**, 267–273 (2005).
49. Kim, H. & Muller, W. J. The role of the epidermal growth factor receptor family in mammary tumorigenesis and metastasis. *Exp. Cell Res.* **253**, 78–87 (1999).
50. Early, E. et al. Transgenic expression of PML/RAR α impairs myelopoiesis. *Proc. Natl Acad. Sci. USA* **93**, 7900–7904 (1996).
51. Park, D. J., Vuong, P. T., de Vos, S., Douer, D. & Koeffler, H. P. Comparative analysis of genes regulated by PML/RAR α and PLZF/RAR α in response to retinoic acid using oligonucleotide arrays. *Blood* **102**, 3727–3736 (2003).
52. Lewis, J. S. et al. Intrinsic mechanism of estradiol-induced apoptosis in breast cancer cells resistant to estrogen deprivation. *J. Natl Cancer Inst.* **97**, 1746–1759 (2005).
53. Lu, M., Mira-y-Lopez, R., Nakajo, S., Nakaya, K. & Jing, Y. Expression of estrogen receptor α , retinoic acid receptor α and cellular retinoic acid binding protein II genes is coordinately regulated in human breast cancer cells. *Oncogene* **24**, 4362–4269 (2005).
54. Yue, W. et al. Tamoxifen versus aromatase inhibitors for breast cancer prevention. *Clin. Cancer Res.* **11**, 9225s–9230s (2005).
55. Nair, H. B. et al. Induction of aromatase expression in cervical carcinomas: effects of endogenous estrogen on cervical cancer cell proliferation. *Cancer Res.* **65**, 11164–11173 (2005).
56. Ratliff, T. L. Mutation of the androgen receptor causes oncogenic transformation of the prostate. *J. Urol.* **174**, 1149 (2005).
57. Tao, J. et al. Inhibiting the growth of malignant melanoma by blocking the expression of vascular endothelial growth factor using an RNA interference approach. *Br. J. Dermatol.* **153**, 715–724 (2005).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to C.L. (christoph.lengauer@novartis.com).

Mechanisms of drug inhibition of signalling molecules

Judith S. Sebolt-Leopold¹ & Jessie M. English²

The emergence of tumour-specific, molecularly targeted agents signifies a paradigm shift in cancer therapy, with less reliance on drugs that non-discriminately kill tumour and host cells. Although the diversity of targets giving rise to this new generation of anticancer drugs has expanded, many challenges persist in the design of effective treatment regimens. The complex interplay of signal-transduction pathways further complicates the customization of cancer treatments to target single mechanisms. However, despite uncertainty over precise or dominant mechanisms of action, especially for compounds targeting multiple gene products, emerging agents are producing significant therapeutic advances against a broad range of human cancers.

The elucidation of the signal-transduction network that drives neoplastic transformation has led to rationally designed cancer therapeutics that target specific molecular events. Over the past 20 years, we have witnessed an explosion in the launch of drug discovery programmes that, if successful, will significantly lessen our reliance on DNA-directed chemotherapeutic agents with low therapeutic indices as the mainstay of cancer treatment. However, targeted cancer drug candidates, independent of their mechanism of action, frequently face similar hurdles to those that challenge traditional agents. Problems of insufficient efficacy, development of resistance and unacceptable safety profiles continue to hamper clinical progress.

The successful development of molecularly targeted agents is all about making the right choices. Starting with initial target selection, the drug discovery process is fraught with critical junctures at which decisions are difficult. In addition, the stakes are high — each decision must be correct or success will remain elusive, or at best delayed. It all begins with target selection and betting on the right pathway (see the review in this issue by Lengauer and colleagues, page 451). Selection of the appropriate screening strategy for identifying inhibitors/activators around a given target can also be problematic — only to be followed by difficult choices over which ‘hit’ or chemical series a research team should focus their efforts on. As a drug discovery programme advances, preclinical model selection is often not straightforward. The predictability of animal models is a hotly debated topic with no imminent resolution^{1,2}. The cycle of tough choices begins anew with the decision to advance a compound into the clinic. At this point, oncologists and trial sponsors must continue to make the right call on complex issues encompassing patient selection, treatment-regimen design, and the selection of appropriate biomarkers that measure patient response and address proof of concept, often for both the target and the compound selected for development. In the face of these formidable challenges, a number of agents are showing potential. Consequently, we seem to be on the right path in our attempts to treat cancer by disrupting the fundamental signalling pathways that tumours rely on to grow and survive. Here we describe rational points of pharmacological intervention in key signalling pathways. With a focus on kinase targets, this review is intended to provide mechanistic insights and to highlight special considerations that are inherent in the

development of small-molecule inhibitors from both multi-targeted and truly selective classes of agents.

Exploiting the kinome

The *c-src* proto-oncogene was the first to be reported. Its discovery by Bishop and Varmus in 1976 (ref. 3) led to the finding that a 60-kDa phosphoprotein was the long-sought *src* gene product⁴. Shortly thereafter, two laboratories independently demonstrated that the Src protein was a kinase^{5,6}. We now know that there are more than 50 different oncogenes, many of which are protein kinases⁷. The interconnecting role of proto-oncogenes in signalling cascades is finely tuned to maintain the growth and proliferation of normal cells. Upon oncogenic mutation or activation, aberrant cellular signalling drives subsequent tumorigenesis.

Among the first kinases to be identified and linked to tumorigenesis were receptor tyrosine kinases (RTKs). Upon binding of growth factor ligands, RTKs form both homodimers and heterodimers, resulting in activation of their intracellular kinase domains, and subsequent activation of downstream signalling cascades that stimulate proliferation and survival (Fig. 1). Agents targeting RTKs in oncology include therapeutic antibodies (that is, biological agents or biologics) to RTK ligands or the receptors themselves, and small-molecule inhibitors that target the intracellular kinase domains of the RTKs. Among the most aggressively targeted RTKs are the ErbB and vascular endothelial growth factor receptor (VEGFR) families of tyrosine kinases^{8,9}. As depicted in Fig. 1, two major intracellular signalling cascades that are activated by tyrosine kinase receptors and co-opted in tumour cells are the Ras–mitogen-activated protein kinase (MAPK) and phosphatidylinositol-3-OH kinase (PI(3)K)–AKT–mTOR (mammalian target of rapamycin) pathways (see the review in this issue by Shaw and Cantley, page 424).

The entire collection of kinases encoded by the human genome, which is known as the kinome and encompasses over 500 protein kinases, offers a rich and diverse source of potentially druggable targets for disrupting tumour growth and survival. Eight kinase-targeted oncology drugs have received regulatory approval so far (Table 1), and more than 100 additional agents are currently undergoing clinical evaluation. The diversity of kinases being targeted by these test small-molecule

¹Pfizer Global Research and Development, Michigan Laboratories, 2800 Plymouth Road, Ann Arbor, Michigan 48105, USA. ²Pfizer Research Technology Center, 620 Memorial Drive, Cambridge, Massachusetts 02139, USA.

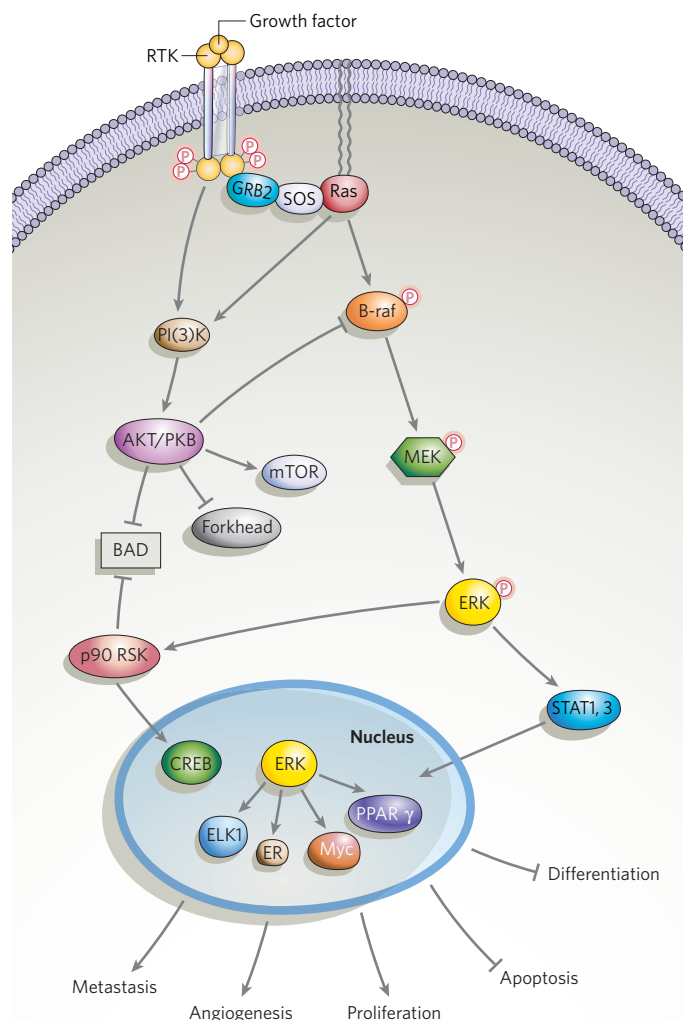


Figure 1 | Central growth factor signalling pathways that drive pleiotropic cellular responses. In tumour cells, a number of receptor tyrosine kinases (RTKs), which include the ErbB and vascular endothelial growth factor receptor families of tyrosine kinases, can become activated by various mechanisms, including mutation, overexpression and autocrine or paracrine production of their respective growth factor-family ligands. In response to ligand-induced activation of cell-surface RTKs, assembly complexes are formed that contain activated, autophosphorylated growth factor receptors with adaptor proteins such as growth factor receptor-bound-2 (GRB2) and exchange factors such as Son-of-sevenless (SOS), to activate Ras. Adaptor proteins are composed of Src homology 2 (SH2) and SH3 domains, which serve as docking sites for various signalling proteins, including receptors and GTPase regulators. In cooperation with GRB2, SOS activates Ras by catalysing the replacement of GDP with GTP. In its GTP-bound form, Ras initiates membrane recruitment and activation of Raf, leading to the activation of the dual specificity mitogen-activated protein kinase (MAPK) kinases MEK1 and MEK2. Activated MEK continues the cascade of phosphorylation events by activating MAPK/extracellular signal-regulated kinase-1 (ERK1) and ERK2. Once it is phosphorylated, ERK is translocated across the nuclear membrane, resulting in the activation of numerous transcription factors, including members of the ETS family that regulate cell-cycle progression (ELK1), as well as oestrogen receptor (ER), Myc, c-Fos, peroxisome proliferator-activated receptor- γ (PPAR- γ), and signal transducer and activator of transcription 1 (STAT1) and STAT3. Activated ERK also phosphorylates cytoplasmic p90 ribosomal protein S6 kinase (RSK), leading to phosphorylation and inactivation of the pro-apoptotic protein BAD. RSK activation further promotes cell survival by leading to the phosphorylation of cyclic-AMP-responsive element-binding protein transcription factor (CREB). Ras has a central role in mediating not only proliferation signalling through the MAPK pathway, but also survival signalling through the phosphatidylinositol-3-OH kinase (PI(3)K) pathway. Activated Ras interacts with PI(3)K to generate second-messenger lipids that are critical for activation of numerous target proteins, including the survival signalling kinase AKT/protein kinase B (PKB). AKT provides strong anti-apoptotic signals through its negative regulation of Raf, forkhead transcription factors and BAD. The PI(3)K-AKT pathway is also important in modulating mammalian target of rapamycin (mTOR), which is a serine/threonine kinase that acts as a central sensor for nutrient/energy availability, thereby regulating cell growth in response to the environment.

compounds or biologics is shown in Table 2. Of those agents in late-stage development, the majority target tyrosine protein kinases.

Modes of small-molecule inhibitor binding

The general architecture of a protein kinase is depicted in Fig. 2. Protein kinases possess a deep ATP-binding active site that also affords a high-affinity binding site for small-molecule inhibitors. Thus, many of the first kinase inhibitors were ATP mimetics that bound to this site and competed with cellular ATP. As the ATP site is highly conserved across the kinome, dogma suggested that it would be difficult to identify selective inhibitors that targeted only the therapeutically relevant kinase. It was feared that kinase inhibitors would be non-selective, thereby precluding the realization of a sufficient therapeutic index. Early evidence suggesting a path forward was provided by studies on isoforms of the p38 MAPK, indicating that residues close to the ATP-binding site could be exploited to achieve selectivity, even among closely related kinases^{10,11}. Refinement continues with strategies that exploit amino-acid heterogeneity in, or close to, the ATP-binding site of kinases. Key to these efforts has been the elucidation of the human kinome combined with increasing numbers of solved kinase structures^{12,13}. Most of the 75 or more kinase inhibitors that are currently in clinical trials seem to be ATP mimetic. However, the discovery of alternative binding modes for small-molecule inhibitors (Fig. 3) is being exploited to build a high degree of selectivity into drug molecules.

The first non-classical kinase inhibitors to be identified were the MAPK kinase (MEK) inhibitors PD98059 (ref. 14) and U0126 (ref. 15). It was recognized early on that these inhibitors were non-ATP-competitive and probably interacted with MEK1 and MEK2 in a unique way.

This novel binding mode seemed to confer desirable properties, such as enhanced selectivity relative to ATP-mimetic inhibitors¹⁶. Recently, this unusual binding mode was revealed in co-crystal structures demonstrating simultaneous binding of ATP and a close structural analogue of the MEK inhibitor CI-1040, confirming that these inhibitors are not ATP competitive¹⁷. The occupation of a unique inhibitor-binding pocket adjacent to the ATP site by CI-1040-like MEK inhibitors is thought to induce several conformational changes in unphosphorylated MEK, consequently serving to lock the enzyme in a closed but catalytically inactive form. Notably, the MEK inhibitor-binding pocket is located in a region where sequence similarity with other protein kinases is relatively low and distinct from the homologous ATP-binding site¹⁷. Therefore, nature has cooperated in providing drug researchers with an exploitable mechanism for developing highly selective inhibitors that block MEK-catalysed activation of the MAPK pathway.

As our database of small-molecule kinase inhibitors has expanded, additional exploitable binding modes have been revealed. Birb796, imatinib (Gleevec; STI571) and sorafenib (Nexavar; BAY 43-9006) bind in similar ways to the active site of p38 α , Abl and B-raf kinases, respectively, extending beyond the highly conserved ATP-binding site¹⁸⁻²⁰ (see Fig. 3 for a depiction of Birb796 binding to p38). The elucidation of co-crystal structures of imatinib analogues with c-Abl, and Birb796 with p38, showed that these inhibitors take advantage of the conformational plasticity of kinases to constrain them in an inactive state. Whereas there are probably few conformations that a kinase can adopt to be active, they seem to have a range of conformations in the inactive state. Thus, the enhanced selectivity of these inhibitors derives, in part, from their binding to the inactive conformation of the kinase. The

Table 1 | Clinically approved kinase-targeted oncology agents

Drug	Number	Known targets
Small molecules		
Imatinib (Gleevec)	STI-571	Abl, PDGFR, c-Kit
Gefitinib (Iressa)	ZD-1839	EGFR
Erlotinib (Tarceva)	OSI-774/CP358774	EGFR
Sorafenib (Nexavar)	BAY 43-9006	VEGFR, PDGFR, FLT3, c-Kit, B-raf, Raf-1
Sunitinib (Sutent)	SU11248	VEGFR, PDGFR, FLT3, c-Kit
Biologics		
Trastuzumab (Herceptin)	–	ErbB2 (HER-2/neu)
Cetuximab (Erbix)	–	EGFR
Bevacizumab (Avastin)	–	VEGF

Agents included in the table are approved by the Food and Drug Administration for use in the United States. Abl, Abelson leukaemia virus; EGFR, epidermal growth factor receptor; PDGFR, platelet-derived growth factor receptor; VEGFR, vascular endothelial growth factor receptor.

selectivity profile of these molecules cannot be predicted on the basis of highly conserved sequence similarity alone. Methods for predicting inhibitor cross-reactivity profiles are being developed²¹. As we enhance our understanding of the conformational plasticity of kinases, our ability to identify novel inhibitors that target the back pocket of kinases will be facilitated. In so doing, we will be able to further refine the selectivity profiles of target drug molecules.

Multi-targeted kinase inhibitors

As most cancers are the result of a number of mutations²², it is reasonable to expect agents that target a number of different kinases to have a better chance of efficacy than highly selective kinase inhibitors administered as single agents. For decades, oncologists have taken advantage of combinations of agents with unique activities for the management of neoplastic disease. Combination therapy is the norm rather than the exception. However, it is difficult to rationally design small-molecule inhibitors that are highly potent and selective against a pre-determined array of desired kinase targets unless these targets are structurally highly similar. The discovery of multi-targeted kinase inhibitors has been largely empirical in the sense that many have evolved from drug discovery programmes in which non-selective ATP-competitive chemical matter was identified at the outset. Our understanding of cellular signalling and tumour genetics is not currently refined to the point that predictions can be successfully made regarding which combinations of kinase inhibition will yield maximal efficacy in a given tumour type. Therefore, most multi-targeted discovery programmes have two key laboratory objectives: first, to optimize activity against the critical targets; and second, to address the question of whether accompanying activity against another target is a positive attribute or a liability that needs to be designed out. Here we focus on two clinically successful multi-targeted kinase inhibitors — sunitinib (Sutent; SU11248) and sorafenib (Nexavar; BAY 43-9006) — to illustrate how distinctly different preclinical approaches led to approved agents with proven clinical efficacy against renal tumours.

Sunitinib emerged from a Sugen research programme that, from its early days, was focused on targeting the split kinase domain superfamily of RTKs. These proteins, encompassing VEGFR family members as well as platelet-derived growth factor receptor- β (PDGFR- β), Kit and FLT3, are expressed on solid tumour cells, and participate in autocrine loops implicated in cancer growth and survival^{23–25}. In addition, several of the split kinase domain RTKs, namely the VEGFRs and PDGFR- β , are key angiogenesis targets²⁶. Thus, a multi-targeted kinase inhibitor capable of inhibiting all or several members of this family could potentially result in broad-spectrum anti-tumour efficacy. The subsequent demonstration that sunitinib possessed significant *in vivo* activity against a diverse panel of human xenografts provided further impetus for testing this hypothesis in the clinic²⁷. The integration of biomarker readouts into these preclinical studies as biological indicators of drug function proved important in guiding the early clinical evaluation of this agent²⁷. Pharmacokinetics

(PK; the time and dose dependence of the plasma drug concentration) were analysed relative to Sutent's pharmacodynamics (PD; the time and dose dependence of target modulation) and efficacy. By determining the PK–PD relationship, critical proof-of-concept criteria for a given drug candidate can be established before proceeding into early clinical trials. Importantly, an increase in VEGF-A and PDGF, and a concurrent decrease in sVEGFR-2, were subsequently shown in plasma samples analysed from patients treated with sunitinib²⁸. sVEGFR-2, which is a soluble form of VEGFR-2, similar to sVEGFR-1 and other soluble circulating RTKs, is a potential quantitative biomarker of angiogenesis and anti-angiogenic drug activity²⁹. Importantly, 40% of renal cancer patients treated with sunitinib were classified as partial responders, and the overall progression-free survival (PFS) of all patients under study was 8.7 months²⁸. This robust clinical activity of sunitinib against metastatic renal cell carcinoma led to its recent regulatory approval for this indication. In addition, sunitinib has been approved for treatment of imatinib-refractory gastrointestinal stromal tumours (GISTs).

Sorafenib is also a multi-targeted kinase inhibitor, which was recently approved for treatment of metastatic renal cancer. However, in the early days of its preclinical development, the ability of sorafenib to inhibit VEGFR and angiogenesis was not fully appreciated. Early reports on sorafenib were exclusively devoted to accounts of its targeting of Raf-1, presumably by virtue of its inhibition of Raf kinase, which is a serine/threonine kinase central to the MAPK pathway^{30,31}. A putative Raf-directed mechanism of action was consistent with sorafenib having been identified from high-throughput screening of small molecules against c-raf kinase. Crystallographic studies showed that sorafenib binds to the ATP pocket of B-raf, interacting with residues in both the P-loop and the kinase-activation loop. It is believed that inhibition of Raf catalytic activity by sorafenib is achieved by its ability to prevent the activation loop and the catalytic residues from adopting a conformation that is competent to bind and phosphorylate substrate²⁰. After this agent entered clinical evaluation, the multi-targeted nature of sorafenib became widely known. As reported by Wilhelm and colleagues, this agent was also found to inhibit a number of RTKs involved in angiogenesis³². Most notably, it was shown to inhibit VEGFR-2, VEGFR-3 and PDGFR- β with roughly the same potency as wild-type and mutant B-raf protein kinases. Interestingly, it was also determined that sorafenib inhibits FLT3 and c-Kit³². Based on the clinical activity profile previously discussed for sunitinib, it is therefore not surprising that sorafenib also exhibited renal cancer activity leading to its regulatory approval for this indication. As summarized in the package insert for sorafenib, a 2.8-month prolongation in PFS (167 days versus 84 days for sorafenib-treated and placebo groups, respectively) was observed in a randomized phase III trial of renal cancer patients. The gain in PFS in sorafenib-treated patients primarily reflects the stable disease population, as the response rate was low (2%). Comparative biomarker data are needed to address whether the

Table 2 | The pipeline: representative kinase-directed oncology drug targets

Kinome branch	Compound	Known targets
TK	AZD-2171	VEGFR
	VandetanibZD6474	VEGFR, EGFR
	Vatalanib/PTK787	VEGFR, PDGFR, c-Kit
	Axitinib/AG-013736	VEGFR, PDGFR
	Lapatinib/GW572016	EGFR, ErbB2
	Dasatinib/BMS-354825	Src, Abl
CMGC	AMN-107	Abl
	CP-751871	Anti-IGFR antibody
	Flavopiridol	Pan-cdk
STE	PD332991	cdk4
	PD0325901	MEK
	ARRY-142886	MEK

Abl, Abelson leukaemia virus; cdk4, cyclin-dependent kinase-4; CMGC, cyclin-dependent kinase/mitogen-activated protein kinase/glycogen synthase kinase/cdk-like kinase; EGFR, epidermal growth factor receptor; IGFR, insulin-like growth factor receptor; MEK, mitogen-activated protein kinase; PDGFR, platelet-derived growth factor receptor; STE, sterol; TK, tyrosine kinase; VEGFR, vascular endothelial growth factor receptor.

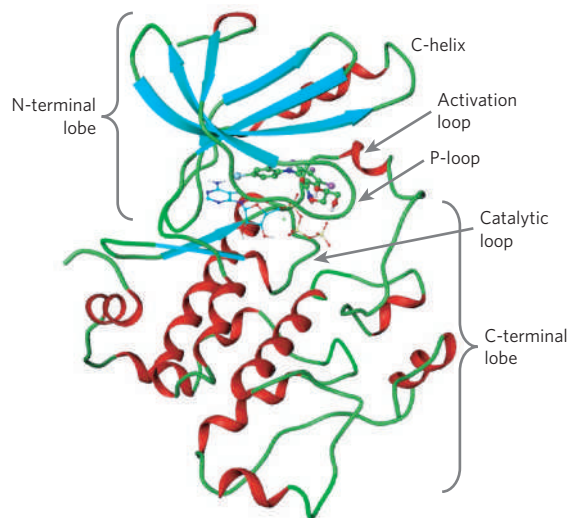


Figure 2 | The protein kinase fold: ribbon diagram of MEK1 with bound ATP and PD318088. The architecture of mitogen-activated protein kinase 1 (MEK1) is shown with the MEK1 inhibitor PD318088 and ATP bound simultaneously. Protein kinases form a conserved protein fold with an amino-terminal domain that primarily comprises β -strands, and a carboxy-terminal domain that is primarily α -helical. The ATP-binding and protein substrate-binding sites of protein kinases reside at the interface of these two domains. The highly conserved structural elements that are involved in regulating activity and catalysis are labelled. For many protein kinases, phosphorylation of the activation loop results in activation of the kinase and large conformational changes in the activation loop^{57,58}. The P-loop is involved in alignment and binding of the phosphoryl groups of ATP. The proper alignment of the C-helix is required for productive binding of ATP and orientation of the active site into a productive mode. The C-helix is often misaligned in inactive kinases. The catalytic loop contains highly conserved residues required for phosphotransfer.

decreased response rate of sorafenib relative to sunitinib reflects differences in potency against VEGFR family members. It is unclear whether treatment with sorafenib results in clinically meaningful efficacy that is directly attributable to its inhibition of Raf kinase. As clinical trial data with sorafenib mature, it will be important to probe how well B-raf-mutated tumours respond to this agent.

It is generally difficult to ascertain the contribution of individual target activities when a multi-targeted kinase inhibitor shows clinical efficacy. However, preclinical studies can often provide clues. For example, studies reported by Bergers and colleagues³³ demonstrated that the combination of a VEGFR inhibitor that was structurally related to sunitinib, and imatinib, which inhibited PDGFR signalling, resulted in superior activity compared with VEGFR inhibitor alone. The two inhibitors target both tumour-associated pericytes via PDGFR, and endothelial cells via VEGFR³³. Thus, the activity of sunitinib and sorafenib against renal cancer might be due, in part, to the ability of these agents to target both of these components of the tumour vasculature.

Highly selective kinase inhibitors

Antibodies, by virtue of the fact that they hit single targets, represent a rational approach for obtaining highly selective agents. The first protein kinase inhibitor to be approved was the monoclonal antibody trastuzumab (Herceptin), which targets the ErbB2 (HER-2/neu) receptor. This agent has subsequently become an important component of therapy for HER-2-positive breast cancer³⁴. The monoclonal antibodies cetuximab (Erbix) and bevacizumab (Avastin), targeting epidermal growth factor receptor (EGFR) and VEGF, respectively, have also gained regulatory approval^{35,36}. Many more antibody-based clinical candidates are currently in clinical trials.

Another approach to the design of highly selective kinase inhibitors has focused on small molecules. Despite widespread scepticism, the ability to design highly selective ATP-mimetic kinase inhibitors has

now been demonstrated for a number of targets. Drug discovery programmes can take considerably longer when faced with the laboratory objective of designing out contaminating kinase activity. However, if successful, such programmes produce clean (that is, specific) inhibitors that serve as useful tools for probing the role of a given target in driving the phenotype of individual tumours. Consequently, such agents are potentially amenable to providing customized therapy to cancer patients. The approved agents gefitinib (Iressa; ZD-1839) and erlotinib (Tarceva; OSI-774) are examples of highly selective ATP-competitive kinase inhibitors targeting EGFR^{37,38}. Other examples include lapatinib, which has dual affinity for ErbB2 and EGFR, and the highly selective cyclin-dependent kinase-4 (cdk4) inhibitor PD0332991, both of which are currently the subject of clinical trials^{39,40}.

Historical screening programmes directed towards protein kinase targets have often used recombinant catalytic kinase domains. It is therefore hardly surprising that the majority of small-molecule inhibitors reported so far are ATP-competitive in nature. The tendency to triage a large collection of hits on the basis of potency alone adds to the likelihood of starting out with highly promiscuous chemical matter. As discussed previously, exceptions are non-classical MEK inhibitors, which are highly selective by virtue of their binding to a unique pocket that is adjacent to, but distinct from, the ATP-binding site¹⁷ (Fig. 3).

Truly selective kinase inhibitors can serve as useful tools for probing the role of a given target and its pathway in various cellular events. The MEK inhibitor PD98059, for example, was released to the academic community, where its usefulness is reflected by its citation in more than 4,000 publications. Single-targeted agents, encompassing both monoclonal antibodies and small-molecule inhibitors, are also potentially useful for the design of combination treatments specifically tailored to known genetic defects of a given tumour. The high incidence of B-raf mutations in melanomas^{41,42}, for example, suggests that they would be highly susceptible to treatment with MAPK pathway inhibitors — for example, MEK or B-raf inhibitors. This prediction was borne out preclinically, as reported by Solit and colleagues for B-raf-mutated melanoma xenografts treated with the MEK inhibitor PD0325901 (ref. 43). However, multiple genetic defects and tumour heterogeneity might make it unlikely that monotherapy with highly selective agents will be sufficient to eradicate tumour burden. The outcome of current clinical testing of the MEK inhibitors PD0325901 and ARRY-142886 (refs 44, 45) in melanoma patients should enhance our understanding of whether this potential challenge translates into a real concern.

Resistance to kinase inhibition

The emergence of drug resistance in treated patients has become a significant issue, and occurs in response to treatment with many approved agents, including imatinib, gefitinib and erlotinib. In some

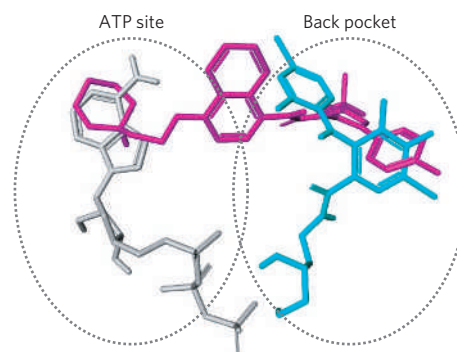


Figure 3 | Alternative binding modes of kinase inhibitors. The binding modes of two kinase inhibitors are shown in relation to the binding site of ATP (grey) in the kinase active site. The MEK1 inhibitor PD318088 (cyan) binds simultaneously with ATP in a region of the kinase active site that is adjacent to the ATP-binding site. Birb796 binding to p38 (purple) extends into the ATP site, but also accesses this back pocket, partly overlapping the region where PD318088 binds in MEK1.

cases, amplification of the oncogenic protein kinase gene can confer resistance⁴⁶. Frequently, however, mutations occur in the kinase catalytic domain, and confer resistance by directly preventing or weakening interaction of the protein with the drug. It is not uncommon to find that drug-resistant mutations in breakpoint cluster region (Bcr)–Abelson leukaemia virus (Abl), Kit and EGFR, are structurally related by virtue of homologous mutations of a conserved ‘gatekeeper’ threonine residue^{47,48}. The clustering of point mutations near protein kinase active sites has led to the search for new therapies against mutated targets resistant to first-line inhibitors. Certain drug-resistant mutants are not necessarily refractory to inhibition with an agent directed against the ATP site of the protein. Examples include the dual Src/Abl kinase inhibitor dasatinib (BMS-354825) and the imatinib derivative AMN107, both of which are effective against imatinib-resistant Abl kinase⁴⁹. Another approach for discovering new agents that are effective against resistant tumours is the employment of alternative *de novo* screening approaches that are less likely to deliver traditional ATP-competitive chemical leads⁵⁰. However, resistance to non-ATP-competitive kinase inhibitors, as observed with the MEK inhibitor CI-1040, can develop in the absence of any mutational changes to the protein, presumably due to changes in expression of other key pathway regulatory molecules⁵¹. It is therefore critical that we anticipate the development of drug resistance when designing novel combination-therapy strategies. For example, *a priori*, the combination of two highly selective agents acting within the same signalling pathway would probably not offer any therapeutic benefit over monotherapy with either agent. However, on a longer-term basis, an advantage might be seen, as manifested by a decreased incidence of resistance to pathway inhibition upon combination of the two agents. In the case of a multi-targeted kinase inhibitor, which can be viewed as combination therapy in a single pill, the development of resistance might be delayed if inhibition of multiple targets contributes to its anti-tumour effects.

Patient selection and design of rational drug combinations

Molecularly targeted therapy will have truly arrived when patients are matched with a treatment regimen predicated by their genetic attributes, as opposed to the histological classification of their malignancies. Gene-expression signatures might prove powerful in determining the extent to which multiple signalling pathways are activated⁵². The design of combination regimens customized to the tumours of individual patients is clearly the correct course and the ultimate goal of current research. As the Ras–MAPK and PI(3)K pathways are strongly interconnected, disruption of one will, in many cases, push tumour cells to increase flux through the other in a virtual tug of war between proliferation and survival signals. Combined inhibition of EGFR and PI(3)K signalling in phosphatase and tensin homologue (PTEN)-mutated tumours has been shown to be a rational and therapeutically advantageous approach⁵³. A number of rational drug combinations can be envisioned that address tumour heterogeneity by targeting different processes or signalling pathways within tumours. There is a common misperception that the individual agents employed in combination trials must first be approved before clinical testing can proceed. However, phase I/II studies of bevacizumab tested in combination with erlotinib for the treatment of non-small-cell lung cancer and renal cancer were conducted prior to the approval of either agent^{54,55}. While intellectual property issues can hinder expedient combination testing of agents that are not solely under the control of a single sponsor, channels nonetheless exist for proceeding with these critical clinical studies. In addition to this ‘designer cocktail’ approach, we should also continue to explore the efficacy of molecularly targeted signalling agents in combination with conventional chemotherapeutic drugs, especially when there is a strong scientific rationale⁵⁶.

Looking forward

Despite our advances in understanding the molecular events that are central to tumorigenesis, only eight kinase-targeted oncology drugs have been approved since 1998. While these statistics reflect incremental progress, there is reason for guarded optimism, as the number and

quality of agents currently undergoing clinical evaluation are relatively high. Yet success will require persistence. Kinases are centre stage in the dysregulation of tumour growth and survival. Although failed agents might come and go, these important protein targets remain a constant feature that must be disrupted if we are to succeed in developing more effective cancer therapies. Consequently, we should be receptive to the need for multi-targeted, as well as single-targeted, kinase inhibitors. There is plenty of therapeutic opportunity for both, and often the distinction separating these two functional classes is ambiguous. For example, imatinib could be viewed as a multi-targeted agent due to its ability to inhibit Abl kinase, c-Kit and PDGFR. Hence, this agent is efficacious against both chronic myelogenous leukaemia (CML) and GIST. However, from the perspective of a CML cell, imatinib probably acts as a single-targeted agent due to the causative role of Bcr–Abl in driving its malignancy. As exemplified by sorafenib, the development of some agents might start out with the belief that they selectively hit one target, but end with the realization that they target multiple kinases, several of which might be important to activity. Semantics notwithstanding, there is a clear need to assemble a collection of pharmaceutically attractive kinase inhibitors that have the proven ability to influence key pathways and to modulate their respective targets at doses that are clinically achievable. Only then will we be in a position to impair multiple pathways non-empirically by judicious combination of selective agents. With that goal in sight, it is imperative that clinical investigators and trial sponsors alike have confidence in these targets and persevere. ■

1. Johnson, J. I. *et al.* Relationships between drug activity in NCI preclinical *in vitro* and *in vivo* models and early clinical trials. *Br. J. Cancer* **84**, 1424–1431 (2001).
2. Voskoglou-Nomikos, T., Pater, J. L. & Seymour, L. Clinical predictive value of the *in vitro* cell line, human xenograft, and mouse allograft preclinical cancer models. *Clin. Cancer Res.* **9**, 4227–4239 (2003).
3. Stehelen, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**, 170–173 (1976).
4. Brugge, J. S. & Erikson, R. L. Identification of a transformation-specific antigen induced by an avian sarcoma virus. *Nature* **269**, 346–348 (1977).
5. Collett, M. S. & Erikson, R. L. Protein kinase activity associated with the avian sarcoma virus src gene product. *Proc. Natl Acad. Sci. USA* **75**, 2021–2024 (1978).
6. Levinson, A. D., Oppermann, H., Levintow, L., Varmus, H. E. & Bishop, J. M. Evidence that the transforming gene of avian sarcoma virus encodes a protein kinase associated with a phosphoprotein. *Cell* **15**, 561–572 (1978).
7. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* **411**, 355–365 (2001).
8. Baselga, J. & Arteaga, C. L. Critical update and emerging trends in epidermal growth factor receptor targeting in cancer. *J. Clin. Oncol.* **23**, 2445–2459 (2005).
9. Jain, R. K., Duda, D. G., Clark, J. W. & Loeffler, J. S. Lessons from phase III clinical trials on anti-VEGF therapy for cancer. *Nature Clin. Pract. Oncol.* **3**, 24–40 (2006).
10. Gum, R. J. *et al.* Acquisition of sensitivity of stress-activated protein kinases to the p38 inhibitor, SB 203580, by alteration of one or more amino acids within the ATP binding pocket. *J. Biol. Chem.* **273**, 15605–15610 (1998).
11. Wang, Z. *et al.* Structural basis of inhibitor selectivity in MAP kinases. *Structure* **6**, 1117–1128 (1998).
12. Kostich, M. *et al.* Human members of the eukaryotic protein kinase family. *Genome Biol.* **3**, RESEARCH0043 (2002).
13. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
14. Dudley, D. T., Pang, L., Decker, S. J., Bridges, A. J. & Saltiel, A. R. A synthetic inhibitor of the mitogen-activated protein kinase cascade. *Proc. Natl Acad. Sci. USA* **92**, 7686–7689 (1995).
15. Favata, M. F. *et al.* Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J. Biol. Chem.* **273**, 18623–18632 (1998).
16. Bain, J., McLauchlan, H., Elliott, M. & Cohen, P. The specificities of protein kinase inhibitors: an update. *Biochem. J.* **371**, 199–204 (2003).
17. Ohren, J. *et al.* Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nature Struct. Biol.* **11**, 1192–1197 (2004).
18. Schindler, T. *et al.* Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. *Science* **289**, 1938–1942 (2000).
19. Pargellis, C. *et al.* Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nature Struct. Biol.* **9**, 268–272 (2002).
20. Wan, C. *et al.* Mechanism of activation of the RAF–ERK signaling pathway by oncogenic mutations of B-Raf. *Cell* **116**, 855–867 (2004).
21. Sheinerman, F. B., Giraud, E. & Laoui, A. High affinity targets of protein kinase inhibitors have similar residues at the positions energetically important for binding. *J. Mol. Biol.* **352**, 1134–1156 (2005).
22. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **7**, 57–70 (2000).
23. Heinrich, M. C., Blanke, C. D., Druker, B. J. & Corless, C. L. Inhibition of KIT tyrosine kinase activity: a novel molecular approach to the treatment of KIT-positive malignancies. *J. Clin. Oncol.* **20**, 1692–1703 (2002).
24. Shimizu, A. *et al.* The dermatofibrosarcoma protuberans-associated collagen type Ia1/platelet-derived growth factor (PDGF) B-chain fusion gene generates a transforming protein that is processed to functional PDGF-BB. *Cancer Res.* **59**, 3719–3723 (1999).
25. Gilliland, D. G. & Griffin, J. D. Role of FLT3 in leukemia. *Curr. Opin. Hematol.* **9**, 274–281 (2002).

26. Cherrington, J. M., Strawn, L. M. & Shawver, L. K. in *Advances in Cancer Research* (eds Klein, G., VandeWoude, G. F.) 1–38 (Academic, San Diego, 2000).
27. Mendel, D. B. *et al.* In vivo antitumor activity of SU11248, a novel tyrosine kinase inhibitor targeting vascular endothelial growth factor and platelet-derived growth factor receptors: determination of a pharmacokinetic/pharmacodynamic relationship. *Clin. Cancer Res.* **9**, 327–337 (2003).
28. Motzer, R. J. *et al.* Activity of SU11248, a multitargeted inhibitor of vascular endothelial growth factor receptor and platelet-derived growth factor receptor, in patients with metastatic renal cell carcinoma. *J. Clin. Oncol.* **24**, 16–24 (2006).
29. Ebos, J. M. *et al.* A naturally occurring soluble form of vascular endothelial growth factor receptor 2 detected in mouse and human plasma. *Mol. Cancer Res.* **2**, 315–326 (2004).
30. Lyons, J. F., Wilhelm, S., Hibner, B. & Bollag, G. Discovery of a novel Raf kinase inhibitor. *Endocrine Relat. Cancer* **8**, 219–225 (2001).
31. Wilhelm, S. & Chien, D.-S. BAY 43-9006: preclinical data. *Curr. Pharm. Design* **8**, 2255–2257 (2002).
32. Wilhelm, S. M. *et al.* BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis. *Cancer Res.* **64**, 7099–7109 (2004).
33. Bergers, G., Song, S., Meyer-Morse, N., Bergsland, E. & Hanahan, D. Benefits of targeting both pericytes and endothelial cells in the tumor vasculature with kinase inhibitors. *J. Clin. Invest.* **111**, 1287–1295 (2003).
34. Yeon, C. H. & Pegram, M. D. Anti-erbB-2 antibody trastuzumab in the treatment of HER2-amplified breast cancer. *Invest. New Drugs* **23**, 391–409 (2005).
35. Goldberg, R. M. *Cetuximab*. *Nature Rev. Drug Discov.* (suppl. 1), S10–S11 (2005).
36. Kerr, D. G. Targeting angiogenesis in cancer: clinical development of bevacizumab. *Nature Clin. Pract. Oncol.* **1**, 39–43 (2004).
37. Cohen, M. H., Williams, G. A., Shridhara, R., Chen, G. & Pazdur, R. FDA drug approval summary: gefitinib (ZD1839) (Iressa) tablets. *Oncologist* **8**, 303–306 (2003).
38. Johnson, J. R. *et al.* Approval summary for erlotinib for treatment of patients with locally advanced or metastatic non-small cell lung cancer after failure of at least one prior chemotherapy regimen. *Clin. Cancer Res.* **11**, 6414–6421 (2005).
39. Nelson, M. H. & Dolder, C. R. Lapatinib: a novel dual tyrosine kinase inhibitor with activity in solid tumors. *Ann. Pharmacother.* **40**, 261–269 (2006).
40. Fry, D. W. *et al.* Specific inhibition of cyclin-dependent kinase 4/6 by PD 0332991 and associated antitumor activity in human tumor xenografts. *Mol. Cancer Ther.* **3**, 1427–1438 (2004).
41. Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
42. Satyamoorthy, K. *et al.* Constitutive mitogen-activated protein kinase activation in melanoma is mediated by both *BRAF* mutations and autocrine growth factor stimulation. *Cancer Res.* **63**, 756–759 (2003).
43. Solit, D. B. *et al.* *BRAF* mutation predicts sensitivity to MEK inhibition. *Nature* **439**, 358–362 (2006).
44. Sebolt-Leopold, J. S. & Herrera, R. Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nature Rev. Cancer* **4**, 937–947 (2004).
45. Wallace, E. M., Lyssikatos, J. P., Yeh, T., Winkler, J. D. & Koch, K. Progress towards therapeutic small molecule MEK inhibitors for use in cancer therapy. *Curr. Topics Med. Chem.* **5**, 215–229 (2005).
46. Gorre, M. C. *et al.* Clinical resistance to STI-571 cancer therapy caused by *BCR-ABL* gene mutation or amplification. *Science* **293**, 876–880 (2001).
47. Tamborini, E. *et al.* A new mutation in the KIT ATP pocket causes acquired resistance to imatinib in a gastrointestinal stromal tumor patient. *Gastroenterology* **127**, 294–299 (2004).
48. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352**, 786–792 (2005).
49. O'Hare, T., Corbin, A. S. & Druker, B. J. Targeted CML therapy: controlling drug resistance, seeking cure. *Curr. Opin. Genet. Dev.* **16**, 92–99 (2006).
50. Adrian, F. J. *et al.* Allosteric inhibitors of Bcr-abl-dependent cell proliferation. *Nature Chem. Biol.* **2**, 95–102 (2006).
51. Wang, Y. *et al.* A role for K-ras in conferring resistance to the MEK inhibitor, CI-1040. *Neoplasia* **7**, 336–347 (2005).
52. Downward, J. Signatures guide drug choice. *Nature* **439**, 274–275 (2006).
53. She, Q.-B. *et al.* The BAD protein integrates survival signaling by EGFR/MAPK and PI3K/Akt kinase pathways in PTEN-deficient tumor cells. *Cancer Cell* **8**, 287–297 (2005).
54. Herbst, R. S. *et al.* Phase I/II trial evaluating the anti-vascular endothelial growth factor monoclonal antibody bevacizumab in combination with the HER-1/epidermal growth factor receptor tyrosine kinase inhibitor erlotinib for patients with recurrent non-small-cell lung cancer. *J. Clin. Oncol.* **23**, 2544–2555 (2005).
55. Hainsworth, J. D. *et al.* Treatment of metastatic renal cell carcinoma with a combination of bevacizumab and erlotinib. *J. Clin. Oncol.* **23**, 7889–7896 (2005).
56. Sawyers, C. Targeted cancer therapy. *Nature* **432**, 294–297 (2004).
57. Canagarajah, B. J., Khokhlatchev, A., Cobb, M. H. & Goldsmith, E. J. Activation mechanism of the MAP kinase ERK2 by dual phosphorylation. *Cell* **90**, 859–869 (1997).
58. Adams, J. A. Activation loop phosphorylation and catalysis in protein kinases: is there functional evidence for the autoinhibitor model? *Biochemistry* **42**, 601–607 (2003).

Acknowledgements The authors wish to thank D. Leopold, R. Herrera and S. Eck for their thoughtful comments. We would like to express special gratitude to J. Ohren for his generous provision of Figs 2 and 3 as well as helpful feedback.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper at www.nature.com/nature. Correspondence should be addressed to J.S.S.-L. (judith.leopold@pfizer.com).

Long γ -ray bursts and core-collapse supernovae have different environments

A. S. Fruchter¹, A. J. Levan^{1,2,3}, L. Strolger^{1,4}, P. M. Vreeswijk⁵, S. E. Thorsett⁶, D. Bersier^{1,7}, I. Burud^{1,8}, J. M. Castro Cerón^{1,9}, A. J. Castro-Tirado¹⁰, C. Conselice^{11,12}, T. Dahlen¹³, H. C. Ferguson¹, J. P. U. Fynbo⁹, P. M. Garnavich¹⁴, R. A. Gibbons^{1,15}, J. Gorosabel^{1,10}, T. R. Gull¹⁶, J. Hjorth⁹, S. T. Holland¹⁷, C. Kouveliotou¹⁸, Z. Levay¹, M. Livio¹, M. R. Metzger¹⁹, P. E. Nugent²⁰, L. Petro¹, E. Pian²¹, J. E. Rhoads¹, A. G. Riess¹, K. C. Sahu¹, A. Smette⁵, N. R. Tanvir³, R. A. M. J. Wijers²² & S. E. Woosley⁶

When massive stars exhaust their fuel, they collapse and often produce the extraordinarily bright explosions known as core-collapse supernovae. On occasion, this stellar collapse also powers an even more brilliant relativistic explosion known as a long-duration γ -ray burst. One would then expect that these long γ -ray bursts and core-collapse supernovae should be found in similar galactic environments. Here we show that this expectation is wrong. We find that the γ -ray bursts are far more concentrated in the very brightest regions of their host galaxies than are the core-collapse supernovae. Furthermore, the host galaxies of the long γ -ray bursts are significantly fainter and more irregular than the hosts of the core-collapse supernovae. Together these results suggest that long-duration γ -ray bursts are associated with the most extremely massive stars and may be restricted to galaxies of limited chemical evolution. Our results directly imply that long γ -ray bursts are relatively rare in galaxies such as our own Milky Way.

It is an irony of astrophysics that stellar birth is most spectacularly marked by the deaths of massive stars. Massive stars burn brighter and hotter than smaller stars, and exhaust their fuel far more rapidly. Therefore a region of star formation filled with low mass stars still early in their lives, and in some cases still forming, may also host massive stars already collapsing and producing supernovae. Indeed, with the exception of the now famous type Ia supernovae, which have been so successfully used for cosmological studies^{1,2} and which are thought to be formed by the uncontrolled nuclear burning of stellar remnants comparable in mass to the Sun³, all supernovae are thought to be produced by the collapse of massive stars. The collapse of the most massive stars (tens of solar masses) is thought to leave behind either black holes or neutron stars, depending largely on the state of chemical evolution of the material that formed the star, whereas the demise of stars between approximately 8 and 20 solar masses produces only neutron stars⁴.

Gamma-ray bursts (GRBs), like supernovae, are a heterogeneous population. GRBs can be divided into two classes: short, hard bursts, which last between milliseconds and about two seconds and have hard high-energy spectra, and long, soft bursts, which last between two and tens of seconds, and have softer high-energy spectra⁵. Only very recently have a few of the short bursts been well localized, and initial studies of their apparent hosts suggest that these bursts may be

formed by the binary merger of stellar remnants^{6,7}. In contrast, the afterglows of over 80 long GRBs (LGRBs) have been detected in the optical and/or radio parts of the spectrum. And as a result of these detections, it has become clear that LGRBs, like core-collapse supernovae, are related to the deaths of young, massive stars. It is these objects, born of the deaths of massive stars, that we study here.

LGRBs are generally found in extremely blue host galaxies^{8–11} that exhibit strong emission lines^{12,13}, suggesting a significant abundance of young, very massive stars. Furthermore, whereas the light curves of the optical transients associated with LGRBs are often dominated by radiation from the relativistic outflow of the GRB, numerous LGRBs have shown late-time ‘bumps’ in their light curves consistent with the presence of an underlying supernova^{14–16}. In several cases spectroscopic evidence has provided confirmation of the light of a supernova superposed on the optical transient^{17–20}. Indeed, given the large variations in the brightnesses of optical transients and supernovae, and the limited observations on some GRBs, it seems plausible that all LGRBs have an underlying supernova²¹. And although the energy released in a LGRB often appears to the observer to be orders of magnitude larger than that of a supernova, there is now good evidence suggesting that most LGRBs are highly collimated and often illuminate only a few per cent of the sky^{22,23}. When one takes this into account, the energy released in LGRBs more closely

¹Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ²Department of Physics and Astronomy, University of Leicester, University Road, Leicester, LE1 7RH, UK. ³Centre for Astrophysics Research, University of Hertfordshire, College Lane, Hatfield, AL10 9AB, UK. ⁴Physics & Astronomy, TCCW 246, Western Kentucky University, 1 Big Red Way, Bowling Green, Kentucky 42101, USA. ⁵European Southern Observatory, Alonso de Córdova 3107, Casilla 19001, Santiago, Chile.

⁶Department of Astronomy & Astrophysics, University of California, 1156 High Street, Santa Cruz, California 95064, USA. ⁷Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead, CH41 1LD, UK. ⁸Norwegian Meteorological Institute, PO Box 43, Blindern, N-0313 Oslo, Norway. ⁹Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen, Denmark. ¹⁰Instituto de Astrofísica de Andalucía (CSIC), Camino Bajo de Huétor, 50, 18008 Granada, Spain. ¹¹California Institute of Technology, Mail Code 105-24, Pasadena, California 91125, USA. ¹²School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, NG7 2RD, UK. ¹³Department of Physics, Stockholm University, SE-106 91 Stockholm, Sweden. ¹⁴Physics Department, University of Notre Dame, 225 Nieuwland Hall, Notre Dame, Indiana 46556, USA. ¹⁵Vanderbilt University, Department of Physics and Astronomy, 6301 Stevenson Center, Nashville, Tennessee 37235, USA. ¹⁶Code 667, Extraterrestrial Planets and Stellar Astrophysics, Exploration of the Universe Division, ¹⁷Code 660.1, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ¹⁸NASA/Marshall Space Flight Center, VP-62, National Space Science & Technology Center, 320 Sparkman Drive, Huntsville, Alabama 35805, USA. ¹⁹Renaissance Technologies Corporation, 600 Route 25A, East Setauket, New York 11733, USA. ²⁰Lawrence Berkeley National Laboratory, MS 50F-1650, 1 Cyclotron Road, Berkeley, California 94720, USA. ²¹INAF, Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34131 Trieste, Italy. ²²Astronomical Institute ‘Anton Pannekoek’, University of Amsterdam, Kruislaan 403, NL-1098 SJ Amsterdam, The Netherlands.

resembles that of energetic supernovae. However, not all core-collapse supernovae may be candidates for the production of GRBs. The supernovae with good spectroscopic identifications so far associated with GRBs have been type Ic—that is, core-collapse supernovae that show no evidence of hydrogen or helium in their spectra. (Type Ib supernovae, which are often studied together with type Ic, have spectra that are also largely devoid of hydrogen lines but show strong helium features.) A star may therefore need to lose its outer envelope if a GRB is to be able to burn its way through the stellar atmosphere²⁴. Studies that have compared the locations of type Ib/c supernovae with the more numerous type II supernovae (core-collapse supernovae showing hydrogen lines) in local galaxies so far show no differences in either the type of host or the placement of the explosion on the host^{25,26}. This result led the authors of ref. 25 to argue that core-collapse supernovae all come from the same mass range of progenitor stars, but that type Ib/c supernovae may have had their envelopes stripped by interaction with a binary stellar companion. Whether type Ic supernovae come from single stars, or binary stars, or both, it is very likely that only a small fraction of these supernovae produce GRBs²⁷.

Given the common massive stellar origins of core-collapse supernovae and LGRBs, one might expect that their hosts and local environments are quite similar. It has long been argued that core-collapse supernovae should track the blue light in the Universe (the light from massive stars is blue), both in their distribution among galaxies and within their host galaxies themselves. One would expect similar behaviour from LGRBs, and indeed rough evidence for such a correlation has been reported²⁸. Here we use the high resolution available from Hubble Space Telescope (HST) images, and an analytical technique developed by us that is independent of galaxy morphology, to study the correlation between these objects and the light of their hosts. We also compare the sizes, morphologies and brightnesses of the LGRB hosts with those of the supernovae. Our results reveal surprising and substantial differences between the birthplaces of these cosmic explosions. We find that whereas core-collapse supernovae trace the blue light of their hosts, GRBs are far more concentrated on the brightest regions of their hosts. Furthermore, while the hosts of core-collapse supernovae are approximately equally divided between spiral and irregular galaxies, the overwhelming majority of GRBs are on irregulars, even when we

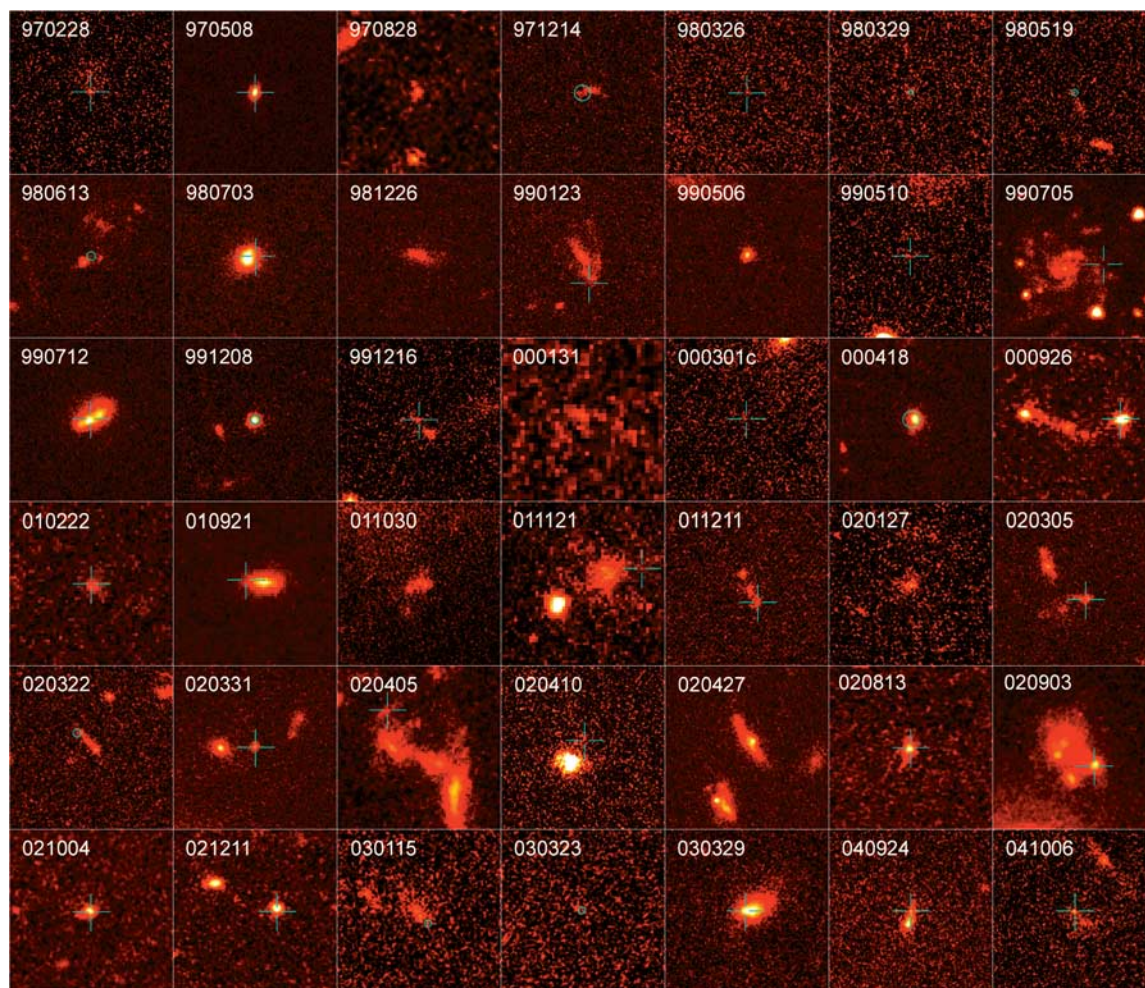


Figure 1 | A mosaic of GRB host galaxies imaged by HST. Each individual image corresponds to a square region on the sky $3.75''$ on a side. These images were taken with the Space Telescope Imaging Spectrograph (STIS), the Wide-Field and Planetary Camera 2 (WFPC2) and the Advanced Camera for Surveys (ACS) on HST. In cases where the location of the GRB on the host is known to better than $0.15''$ the position of the GRB is shown by a green mark. If the positional error is smaller than the point spread function of the image ($0.07''$ for STIS and ACS, $0.13''$ for WFPC2) the position is marked by a cross-hair; otherwise the positional error is indicated by a circle. The STIS images were all taken in white light (no filter), and in most cases

the WFPC2 and ACS images are in the F606W filter (though in a few cases where images in this filter were not available we have used images in F555W or F775W). The STIS and F606W images can be thought of as broad 'V' or visual images, and are, for galaxies exhibiting typical colours of GRB hosts, the single most sensitive settings for these cameras. F555W is close to the ground-based Johnson V band, and F775W corresponds to the ground-based Johnson I band. Owing to the redshifts of the hosts, these images generally correspond to blue or ultraviolet images of the hosts in their rest frame, and thus detect light largely produced by the massive stars in the hosts.

restrict the GRB sample to the same redshift range as the supernova sample. We argue that these results may be best understood if GRBs are formed from the collapse of extremely massive, low-metallicity stars.

The sample

Over 40 LGRBs have been observed with the HST at various times after outburst. The HST is unique in its capability easily to resolve the distant hosts of these objects. Shown in Fig. 1 is a mosaic of HST images of the hosts of 42 bursts. These are all LGRBs with public data that had an afterglow detected with better than 3σ significance and a position sufficiently well localized to determine a host galaxy. A list of all the GRBs used in this work can be found in Supplementary Tables 1–3.

The supernovae discussed here were all discovered as part of the Hubble Higher z Supernova Search^{29,30}, which was done in cooperation with the HST GOODS survey³¹. The GOODS survey observed two ~ 150 arcmin² patches of sky five times each, in epochs separated by 45 days. Supernovae were identified by image subtraction. Here we discuss only the core-collapse supernovae identified in this survey. A list of the supernovae used is presented in Supplementary Table 4, and images of the supernova hosts can be seen in Fig. 2.

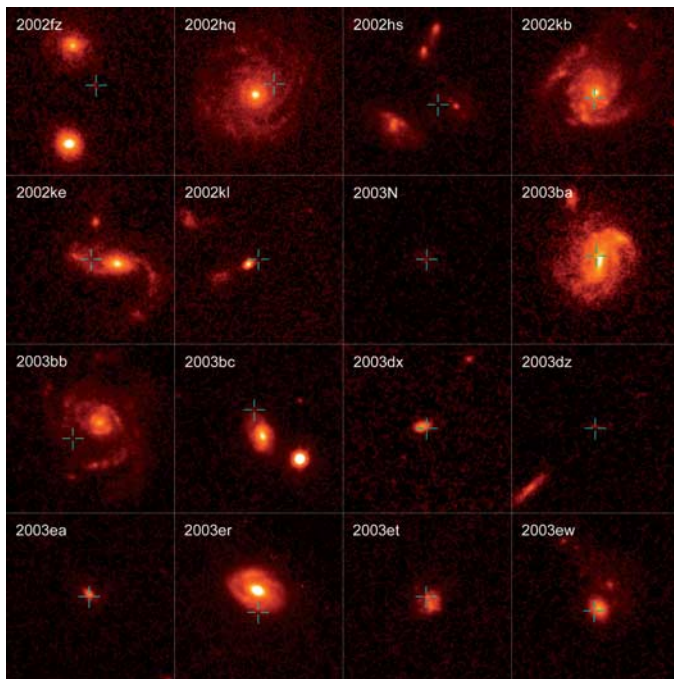


Figure 2 | A mosaic of core-collapse supernova host galaxies imaged with HST as part of the GOODS programme. Each image in the mosaic has a width of $7.5''$ on the sky, and thus twice the field of view of each image in the GRB mosaic. The position of each supernova on its host galaxy is marked. In all cases, these positions are known to sub-pixel accuracy. Supernovae in the GOODS sample were identified by ref. 30 as either type Ia or core-collapse supernovae on the basis of their colours, luminosities and light curves, as data allowed (a supernova exploding near the beginning or end of one of the multi-epoch observing runs would have much less data, and sometimes poor colour information). Thus bright type Ib and Ic supernovae, which have colours and luminosities similar to type Ia supernovae, would probably have been classified as type Ia (unless a grism spectrum was taken—however, only a small fraction of objects were observed spectroscopically). On the other hand, fainter type Ib and Ic supernovae ($M_B \gtrsim -18$) could in principle be identified from photometric data; however, in practice the data were rarely sufficient for a clear separation from other core-collapse supernovae. On the basis of surveys of nearby galaxies, one might expect approximately 20% of the core-collapse supernovae to be type Ib or Ic^{49,50}.

Positions of GRBs and supernovae on their hosts

If LGRBs do in fact trace massive star formation, then in the absence of strong extinction we should find a close correlation between their position on their host galaxies and the blue light of those galaxies. However, many of the GRB hosts and quite a few of the supernova hosts are irregular galaxies made up of more than one bright component. As a result, the common astronomical procedure of identifying the centroid of the galaxy's light, and then determining the distance of the object in question from the centroid, is not particularly appropriate for these galaxies—the centroid of light may in fact lie on a rather faint region of the host (examine GRBs 000926 and 020903 in Fig. 1 for excellent illustrations of this effect). We have therefore developed a method that is independent of galaxy morphology. We sort all of the pixels of the host galaxy image from faintest to brightest, and ask what fraction of the total light of the host is contained in pixels fainter than or equal to the pixel containing the explosion. If the explosions track the distribution of light, then the fraction determined by this method should be uniformly distributed between zero and one. (A detailed exposition of this method can be found in Supplementary Information).

As can be seen in Fig. 3, the core-collapse supernovae do track the light of their hosts as well as could be expected given their small number statistics. A Kolmogorov–Smirnov (KS) test finds that the distribution of the supernovae is indistinguishable from the distribution of the underlying light. The situation is clearly different for LGRBs. As can be seen in Fig. 3, the GRBs do not simply trace the blue light of the hosts; rather, they are far more concentrated on the peaks of light in the hosts than the light itself. A KS test rejects the hypothesis that GRBs are distributed as the light of their hosts with a probability greater than 99.98%. Furthermore, this result is robust:

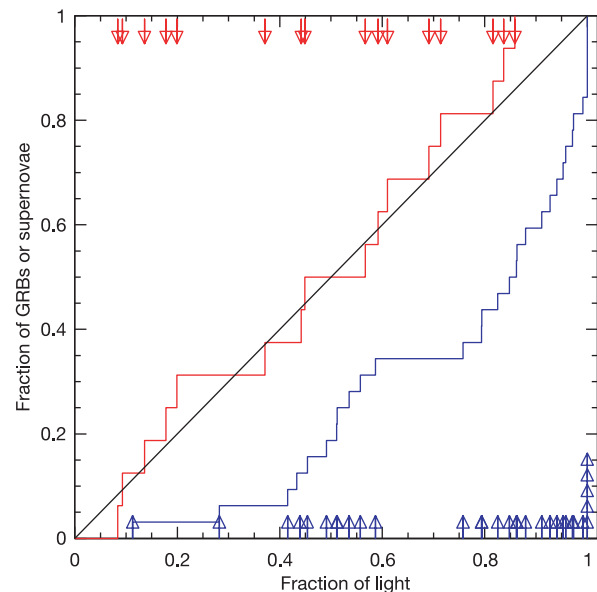


Figure 3 | The locations of the explosions in comparison to the host light. For each object, an arrow indicates the fraction of total host light in pixels fainter than or equal to the light in the pixel at the location of the transient. The cumulative fraction of GRBs or supernovae found at a given fraction of the total light is shown as a histogram. The blue arrows and histogram correspond to the GRBs, and the red arrows and histogram correspond to the supernovae. Were the GRBs and supernovae to track the light identically, their histograms would follow the diagonal line. Whereas the supernova positions do follow the light within the statistical error, the GRBs are far more concentrated on the brightest regions of their hosts. Thus although the probability of a supernova exploding in a particular pixel is roughly proportional to the surface brightness of the galaxy at that pixel, the probability of a GRB at a given location is effectively proportional to a higher power of the local surface brightness.

it shows no dependence on GRB host size or magnitude. And in spite of the relatively small number of supernova hosts for which a comparison can be made, the two populations are found by the KS test to be drawn from different distributions with $\sim 99\%$ certainty. In the next section, we show that the surprising differences in the locations of these objects on the underlying light of their hosts may be due not only to the nature of their progenitor stars but also that of their hosts.

A comparison of the host populations

An examination of the mosaics of the GRB and supernova hosts (Figs 1 and 2) immediately shows a remarkable contrast—only one GRB host in this set of 42 galaxies is a grand-design spiral, whereas nearly half of the supernova hosts are grand-design spirals. One might wonder if this effect is due to a difference in redshift distribution—the core-collapse supernovae discovered by the GOODS collaboration all lie at redshift $z < 1.2$, whereas LGRBs can be found at much larger redshifts where grand-design spirals are rare to non-existent. Yet if we restrict the GRB population to $z < 1.2$ (and thus produce a population with a nearly identical mean and standard deviation in redshift space compared to the GOODS core-collapse supernovae), the situation remains essentially unchanged: only one out of the eighteen GRB hosts is a grand-design spiral. (For a detailed comparison of GRB hosts to field galaxies, rather than the supernova selected galaxies shown here, see ref. 32.)

Were the difference in spiral fraction the only indication of a difference in the host populations, we could not rule out random chance—given the small number statistics, both populations are barely consistent with each other and a spiral fraction of $\sim 25\%$. However, the host populations differ strongly in ways other than morphology.

In Fig. 4 we compare the 80% light radius (r_{80}) and absolute magnitude distributions of the GRB and supernova hosts. Included

in the comparison are all LGRBs with known redshifts $z < 1.2$ at the time of submission and the 16 core-collapse supernovae of GOODS with spectroscopic or photometric redshifts (see the Supplementary Tables for a complete list of the GRBs, supernovae and associated parameters used in this study). The small minority of GRB hosts in this redshift range without HST imaging are compared only in absolute magnitude and not in size. The absolute magnitudes have been derived from the observed photometry using a cosmology of $\Omega_m = 0.27$, $\Lambda = 0.73$ and $H_0 = 71 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and the magnitudes have been corrected for foreground Galactic extinction³³. For a technical discussion of the determination of the magnitude and size of individual objects, see Supplementary Information.

As can be readily seen, the two host populations differ substantially both in their intrinsic magnitudes and sizes. The GRB hosts are fainter and smaller than the supernova hosts. Indeed, KS tests reject the hypothesis that these two populations are drawn from the same population with certainties greater than 98.6% and 99.7% for the magnitude and size distributions, respectively.

Discussion

Although the evidence is now overwhelming that both core-collapse supernovae and LGRBs are formed by the collapse of massive stars, our observations show that the distribution of these objects on their hosts, and the nature of the hosts themselves, are substantially different. How then can this be? We propose here that these surprising findings are the result of the dependence of the probability of GRB formation on the state of the chemical evolution of massive stars in a galaxy.

Even before the association of LGRBs with massive stars had been established, a number of theorists had suggested that these objects could be formed by the collapse of massive stars, which would leave behind rapidly spinning black holes. An accretion disk about the black hole would power the GRB jet. These models, sometimes referred to as ‘hypernovae’ or ‘collapsar’ models, implicitly require very massive stars, as only stars greater than about 18 solar masses form black holes. But in fact it was widely suspected that even more massive stars would be required—if only to provide the required large energies, and to limit the numbers of supernovae progressing to GRBs.

We conclude that LGRBs do indeed form from the most massive stars and that this is the reason that they are even more concentrated on the blue light of their hosts than the light itself. The most massive stars (O stars) are frequently found in large associations. These associations can be extremely bright, and can indeed provide the peak of the light of a galaxy—particularly if that galaxy is a faint, blue irregular, as are the GRB hosts in general. Indeed, a connection of LGRBs with O stars (and perhaps Wolf-Rayet stars) is a natural one—given the strong emission lines (including Ne [III]) seen in many of these hosts^{12,13} and the evidence for possible strong winds off the progenitors of the GRBs seen in absorption in some LGRB spectra^{34,35}.

However, O stars are found in galaxies of all sizes. Indeed, studies of the Magellanic clouds suggest that the initial distribution of masses of stars at formation in these dwarf galaxies is essentially identical to that in our much larger spiral, the Milky Way³⁶. Therefore, a difference in the initial mass function of stars is unlikely to be responsible for the differences between the hosts. We propose that the fundamental differences between the LGRB and supernova host populations is not their size or luminosity, but rather their metallicity, or chemical evolution. Some evidence of this already exists. The hosts of seven LGRBs (GRBs 980425 (P.M.V., personal communication), 990712¹³, 020903³⁷, 030323³⁸, 030329¹⁷, 031203³⁹ and 050730⁴⁰) have measurements of, or limits on, their metallicity, and in all cases the metallicity is less than one-third solar. The small size and low luminosity of the GRB hosts is then a result of the well known correlation between galaxy mass and metallicity (see ref. 41 and references therein).

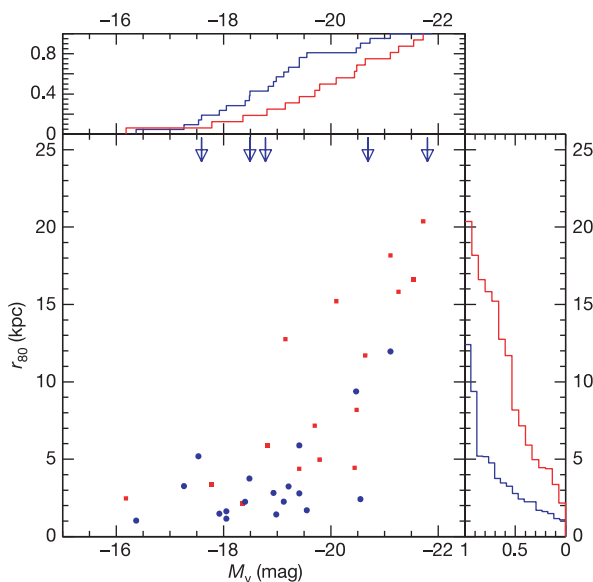


Figure 4 | A comparison of the absolute magnitude and size distributions of the GRB and supernova hosts. In the main panel, the core-collapse supernova hosts are represented as red squares, and the LGRB hosts as blue circles. The absolute magnitudes of the hosts are shown on the x axis, and the lengths of the semi-major axes of the hosts on the y axis. The plot is then projected onto the two side panels where a histogram is displayed for each host population in each of the dimensions—absolute magnitude and semi-major axis. Shown as blue arrows are the absolute magnitudes of GRB hosts with $z < 1.2$ that have been detected from the ground but have not yet been observed by HST. These hosts are only included in the absolute magnitude histogram. The hosts of GRBs are both smaller and fainter than those of supernovae.

But why do LGRBs occur in low-metallicity galaxies? This may be a direct result of the evolution of the most massive stars. It has recently been proposed that metal-rich stars with masses of tens of solar masses have such large winds off their surfaces (due to the photon pressure on their metal-rich atmospheres) that they lose most of their mass before they collapse and produce supernovae⁴. As a result they leave behind neutron stars, not the black holes necessary for LGRB formation. Ironically, stars of 15–30 solar masses may still form black holes, as they do not possess radiation pressure sufficient to drive off their outer envelopes. Direct evidence for this scenario comes from recent work showing that the Galactic soft γ -ray repeater, SGR 1820–06, is in a cluster of extremely young stars of which the most massive have only started to collapse⁴²—yet the progenitor of SGR 1820–06 collapsed to a neutron star, not a black hole. Recent observations of winds from very massive (Wolf-Rayet) stars provide further support for this scenario: outflows from the low-metallicity stars in the Large Magellanic Cloud are substantially smaller than those seen from more metal-rich Galactic stars⁴³. The possible importance of metallicity in LGRB formation has therefore not escaped the notice of theorists^{44,45}.

A preference for low metallicity may also explain one of the most puzzling results of GRB host studies. None of the LGRB hosts is a red, sub-millimetre bright galaxy. These highly dust-enshrouded galaxies at redshifts of ~ 1 –3 are believed to be the sites of a large fraction of the star formation in the distant Universe⁴⁶. And although some LGRB hosts do show sub-millimetre emission, none have the red colours characteristic of the majority of this population. However, it is likely that these red dusty galaxies have substantial metallicities at all redshifts. The low metallicity of hosts may also help explain the fact that a substantial fraction of high-redshift LGRB hosts display strong Lyman- α emission⁴⁷.

All well-classified supernovae associated with LGRBs are of type Ic, presumably because the presence of a hydrogen envelope about the collapsing core can block the emergence of a GRB jet²⁴. Thus only those supernovae whose progenitors have lost some, but not too much, mass appear to be candidates for the formation of a GRB. Given the large numbers of type Ic supernovae in comparison to the estimated numbers of LGRBs, however, it is likely that only a small fraction of type Ic supernovae produce LGRBs. Indeed, even the number of unusually energetic type Ib/Ic supernovae appears to dwarf the LGRB population⁴⁸. Another process, perhaps the spin-up of the progenitor in a binary²⁷, may decide which type Ic supernovae produce LGRBs. Interestingly, it was the similar distribution of supernovae on their hosts, and particularly the fact that type Ib/Ic were no more correlated than type II with the UV bright regions of their hosts, that led ref. 25 to the conclusion that type Ib/Ic supernovae form from binaries. LGRBs clearly track light differently from the general type Ic population. However, the samples used by refs 25 and 26 were from supernovae largely discovered on nearby massive galaxies—dwarf irregular hosts are underrepresented in these samples. It will be particularly interesting to see whether large unbiased supernova surveys at present underway produce similar locations for their supernovae.

We do not know, however, what separates the small fraction of low-metallicity type Ic supernovae that turn into LGRBs from the rest of the population. Potentially, the answer is the amount of angular momentum available in the core to form the jet. In this case, the preference for low metallicity may indicate that single star evolution dominates over binary interaction in forming LGRBs. Deep, high-spectral-resolution studies of LGRB afterglows may provide insight here, by allowing studies of the winds off the progenitor and any binary companion.

Only a small fraction of LGRBs are found in spiral galaxies, even for LGRBs with redshifts $z < 1$ where spirals are much more common. However, the local metallicity in spirals is known to be anti-correlated with distance from the centre of the galaxy. Thus one might expect LGRBs in spirals to violate the trend that we have seen

for the general LGRB population, and avoid the bright central regions of their hosts. The present number of LGRBs known in spirals is still too small to test this prediction. But a sample size a few times larger should begin to allow such a test. Additionally, a survey of the metallicity of the hosts of the GOODS supernovae should find a higher average metallicity than that seen in GRB hosts. Finally, if low metallicity is indeed the primary variable in determining whether LGRBs are produced, then as we observe higher redshifts, where metallicities are lower than in most local galaxies, LGRBs should be more uniformly distributed among star-forming galaxies. Indeed, some evidence of this may already be present in the data³². LGRBs, however, are potentially visible to redshifts as high as $z \approx 10$. At significant redshifts, where the metallicities of even relatively large galaxies are expected to be low, we may find that LGRBs do become nearly unbiased tracers of star formation.

Received 22 August 2005; accepted 5 April 2006.

Published online 10 May 2006.

- Riess, A. G. *et al.* Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116**, 1009–1038 (1998).
- Perlmutter, S. *et al.* Measurements of Ω and Λ from 42 high-redshift supernovae. *Astrophys. J.* **517**, 565–586 (1999).
- Branch, D., Livio, M., Yungelson, L. R., Boffi, F. R. & Baron, E. In search of the progenitors of type Ia supernovae. *Publ. Astron. Soc. Pacif.* **107**, 1019–1028 (1995).
- Heger, A., Fryer, C. L., Woosley, S. E., Langer, N. & Hartmann, D. H. How massive single stars end their life. *Astrophys. J.* **591**, 288–300 (2003).
- Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, 101–104 (1993).
- Gehrels, N. *et al.* A short γ -ray burst apparently associated with an elliptical galaxy at redshift $z = 0.225$. *Nature* **437**, 851–854 (2005).
- Prochaska, J. X. *et al.* The galaxy hosts and large-scale environments of short-hard γ -ray bursts. *Astrophys. J. Lett.* (in the press): preprint at (<http://arxiv.org/astro-ph/0510022>) (2005).
- Fruchter, A. S. *et al.* HST and Palomar imaging of GRB 990123: Implications for the nature of gamma-ray bursts and their hosts. *Astrophys. J.* **519**, 13–16 (1999).
- Sokolov, V. V. *et al.* Host galaxies of gamma-ray bursts: Spectral energy distributions and internal extinction. *Astron. Astrophys.* **372**, 438–455 (2001).
- Le Floc'h, E. *et al.* Are the hosts of gamma-ray bursts sub-luminous and blue galaxies? *Astron. Astrophys.* **400**, 499–510 (2003).
- Christensen, L., Hjorth, J. & Gorosabel, J. UV star-formation rates of GRB host galaxies. *Astron. Astrophys.* **425**, 913–926 (2004).
- Bloom, J. S., Djorgovski, S. G., Kulkarni, S. R. & Frail, D. A. The host galaxy of GRB 970508. *Astrophys. J.* **507**, L25–L28 (1998).
- Vreeswijk, P. M. *et al.* VLT spectroscopy of GRB 990510 and GRB 990712: Probing the faint and bright ends of the gamma-ray burst host galaxy population. *Astrophys. J.* **546**, 672–680 (2001).
- Bloom, J. S. *et al.* The unusual afterglow of the γ -ray burst of 26 March 1998 as evidence for a supernova connection. *Nature* **401**, 453–456 (1999).
- Galama, T. J. *et al.* Evidence for a supernova in reanalyzed optical and near-infrared images of GRB 970228. *Astrophys. J.* **536**, 185–194 (2000).
- Levan, A. *et al.* GRB 020410: A gamma-ray burst afterglow discovered by its supernova light. *Astrophys. J.* **624**, 880–888 (2005).
- Hjorth, J. *et al.* A very energetic supernova associated with the γ -ray burst of 29 March 2003. *Nature* **423**, 847–850 (2003).
- Stanek, K. Z. *et al.* Spectroscopic discovery of the supernova 2003dh associated with GRB 030329. *Astrophys. J.* **591**, L17–L20 (2003).
- Della Valle, M. *et al.* Evidence for supernova signatures in the spectrum of the late-time bump of the optical afterglow of GRB 021211. *Astron. Astrophys.* **406**, L33–L37 (2003).
- Malesani, D. *et al.* SN 2003lw and GRB 031203: A bright supernova for a faint gamma-ray burst. *Astrophys. J.* **609**, L5–L8 (2004).
- Zeh, A., Klose, S. & Hartmann, D. H. A systematic analysis of supernova light in gamma-ray burst afterglows. *Astrophys. J.* **609**, 952–961 (2004).
- Panaiteanu, A. & Kumar, P. Fundamental physical parameters of collimated gamma-ray burst afterglows. *Astrophys. J.* **560**, L49–L53 (2001).
- Frail, D. A. *et al.* Beaming in gamma-ray bursts: Evidence for a standard energy reservoir. *Astrophys. J.* **562**, L55–L58 (2001).
- MacFadyen, A. I., Woosley, S. E. & Heger, A. Supernovae, jets, and collapsars. *Astrophys. J.* **550**, 410–425 (2001).
- van Dyk, S. D., Hamuy, M. & Filippenko, A. V. Supernovae and massive star formation regions. *Astron. J.* **111**, 2017–2027 (1996).
- van den Bergh, S., Li, W. & Filippenko, A. V. Classifications of the host galaxies of supernovae, Set III. *Publ. Astron. Soc. Pacif.* **117**, 773–782 (2005).
- Podsiadlowski, P., Mazzali, P. A., Nomoto, K., Lazzati, D. & Cappellaro, E. The rates of hypernovae and gamma-ray bursts: Implications for their progenitors. *Astrophys. J.* **607**, L17–L20 (2004).

28. Bloom, J. S., Kulkarni, S. R. & Djorgovski, S. G. The observed offset distribution of gamma-ray bursts from their host galaxies: A robust clue to the nature of the progenitors. *Astron. J.* **123**, 1111–1148 (2002).
29. Riess, A. G. *et al.* Identification of Type Ia supernovae at redshift 1.3 and beyond with the Advanced Camera for Surveys on the Hubble Space Telescope. *Astrophys. J.* **600**, L163–L166 (2004).
30. Strolger, L. G. *et al.* The Hubble Higher z Supernova Search: Supernovae to $z \sim 1.6$ and constraints on Type Ia progenitor models. *Astrophys. J.* **613**, 200–223 (2004).
31. Giavalisco, M. *et al.* The Great Observatories Origins Deep Survey: Initial results from optical and near-infrared imaging. *Astrophys. J.* **600**, L93–L98 (2004).
32. Conselice, C. J. *et al.* Gamma-ray burst selected high redshift galaxies: Comparison to field galaxy populations to $z \sim 3$. *Astrophys. J.* **633**, 29–40 (2005).
33. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
34. Mirabal, N. *et al.* GRB 021004: A possible shell nebula around a Wolf-Rayet star gamma-ray burst progenitor. *Astrophys. J.* **595**, 935–949 (2003).
35. Klose, S. *et al.* Probing a gamma-ray burst progenitor at a redshift of $z = 2$: A comprehensive observing campaign of the afterglow of GRB 030226. *Astron. J.* **128**, 1942–1954 (2004).
36. Weidner, C. & Kroupa, P. in *The Initial Mass Function 50 Years Later* (eds Corbelli, E., Pila, F. & Zinnecker, H.) 125–186 (Springer, Dordrecht, 2005).
37. Bersier, D. *et al.* Evidence for a supernova associated with the x-ray flash 020903. *Astrophys. J.* (submitted).
38. Vreeswijk, P. M. *et al.* The host of GRB 030323 at $z = 3.372$: A very high column density DLA system with a low metallicity. *Astron. Astrophys.* **419**, 927–940 (2004).
39. Prochaska, J. X. *et al.* The host galaxy of GRB 031203: Implications of its low metallicity, low redshift, and starburst nature. *Astrophys. J.* **611**, 200–207 (2004).
40. Chen, H.-W., Prochaska, J. X., Bloom, J. S. & Thompson, I. B. Echelle spectroscopy of a GRB afterglow at $z = 3.969$: A new probe of the interstellar and intergalactic media in the young Universe. *Astrophys. J.* **634**, L25–L28 (2005).
41. Kobulnicky, H. A. & Kewley, L. J. Metallicities of galaxies in the GOODS-North Field. *Astrophys. J.* **617**, 240–261 (2004).
42. Figer, D. F., Najarro, F., Geballe, T. R., Blum, R. D. & Kudritzki, R. P. Massive stars in the SGR 1806-20 cluster. *Astrophys. J.* **622**, L49–L52 (2005).
43. Crowther, P. A. & Hadfield, L. J. Reduced Wolf-Rayet line luminosities at low metallicity. *Astron. Astrophys.* **449**, 711–722 (2006).
44. Woosley, S. & Heger, A. The progenitor stars of gamma-ray Bursts. *Astrophys. J.* **637**, 914–921 (2005).
45. Yoon, S.-C. & Langer, N. Evolution of rapidly rotating metal-poor massive stars towards gamma-ray bursts. *Astron. Astrophys.* **443**, 643–648 (2005).
46. Chapman, S. C., Blain, A. W., Smail, I. & Ivison, R. J. A redshift survey of the submillimeter galaxy population. *Astrophys. J.* **622**, 772–796 (2005).
47. Fynbo, J. P. U. *et al.* On the Ly α emission from gamma-ray burst host galaxies: Evidence for low metallicities. *Astron. Astrophys.* **406**, L63–L66 (2003).
48. Soderberg, A. M., Nakar, E., Kulkarni, S. R. & Berger, E. Late-time radio observations of 68 Type Ibc supernovae: Strong constraints on off-axis gamma-ray bursts. *Astrophys. J.* **638**, 930–937 (2006).
49. van den Bergh, S. & Tammann, G. A. Galactic and extragalactic supernova rates. *Annu. Rev. Astron. Astrophys.* **29**, 363–407 (1991).
50. Mannucci, F. *et al.* The supernova rate per unit mass. *Astron. Astrophys.* **433**, 807–814 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Support for this research was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc. Observations analysed in this work were taken by the NASA/ESA Hubble Space Telescope under programmes: 7785, 7863, 7966, 8189, 8588, 9074 and 9405 (Principal Investigator, A.S.F.); 7964, 8688, 9180 and 10135 (PI, S. R. Kulkarni); 8640 (PI, S.T.H.). We thank N. Panagia, N. Walborn and A. Soderberg for conversations; A. Filippenko and collaborators for early-time images of GRB 980326; and J. Bloom and collaborators for making public their early observations of GRB 020322.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.S.F. (fruchter@stsci.edu).

RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse

Minoo Rassoulzadegan^{1,2}, Valérie Grandjean^{1,2}, Pierre Gounon³, Stéphane Vincent^{1,2†}, Isabelle Gillot^{1,2} & François Cuzin^{1,2}

Paramutation is a heritable epigenetic modification induced in plants by cross-talk between allelic loci. Here we report a similar modification of the mouse *Kit* gene in the progeny of heterozygotes with the null mutant *Kit*^{tm1Alf} (a *lacZ* insertion). In spite of a homozygous wild-type genotype, their offspring maintain, to a variable extent, the white spots characteristic of *Kit* mutant animals. Efficiently inherited from either male or female parents, the modified phenotype results from a decrease in *Kit* messenger RNA levels with the accumulation of non-polyadenylated RNA molecules of abnormal sizes. Sustained transcriptional activity at the postmeiotic stages—at which time the gene is normally silent—leads to the accumulation of RNA in spermatozoa. Microinjection into fertilized eggs either of total RNA from *Kit*^{tm1Alf/+} heterozygotes or of *Kit*-specific microRNAs induced a heritable white tail phenotype. Our results identify an unexpected mode of epigenetic inheritance associated with the zygotic transfer of RNA molecules.

Paramutation, first observed in maize¹ and subsequently in a variety of plants², is a heritable epigenetic change in the phenotype of a 'paramutable' allele, initiated by interaction in heterozygotes with a 'paramutagenic' form of the locus. Often referred to as an exception to the law of Mendel, which states that genetic factors segregate unchanged from heterozygotes, paramutation is meiotically stable and inherited in the absence of the inducing allele. So far, the closest observations of paramutation in an animal species are changes in DNA methylation profiles directed by the allelic locus in the mouse, which we and others described as 'transvection' or 'paramutation-like' effects^{3,4}. Here we report a modification in the phenotypic expression of the wild-type allele of the gene encoding the *Kit* receptor in the progeny of heterozygotes with a null insertion mutant. The 'paramutated' (*Kit*^{*}), genotypically wild-type animals maintain the white-spotted phenotype that is characteristic of *Kit* mutants in the absence of the mutant allele. The efficient paternal and maternal inheritance of the paramutated state raises the question of a possible molecular support for the epigenetic information.

Non-mendelian phenotype distribution

The *tm1Alf* mutation (Mouse Genome Informatics accession identifier MGI:2449782; initially designated *Kit*^{W-lacZ}) was engineered⁵ by inserting a 3-kilobase (kb) *lacZ-neo* cassette downstream of the initiator ATG site (Fig. 1a). A unique mRNA of the same size containing the β -galactosidase coding sequence is expressed under the control of the *Kit* promoter and regulatory sequences. The mutation abrogates the synthesis of the *Kit* tyrosine kinase receptor, which has a critical role in several developmental processes including germ cell differentiation, haematopoiesis and melanogenesis. Accordingly, *Kit*^{tm1Alf} homozygotes die shortly after birth, and heterozygotes show a white tail tip and white feet (Fig. 1a). We initially observed an abnormal segregation of phenotypes in the progeny of crosses between two heterozygous parents. Wild-type genotypes were identified by the absence of *lacZ* sequences (determined by genomic polymerase chain reaction (PCR) analysis) and

β -galactosidase expression (determined by *in situ* 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) staining) (data not shown), and further confirmed by Southern blot analysis (Fig. 1c). However, it was notable that most of these genetically wild-type (*Kit*^{+/+}) mice maintained the white patches characteristic of the parental heterozygotes (Fig. 1b and Table 1). The occurrence of this modified 'paramutated' form of the *Kit*⁺ allele (*Kit*^{*} phenotype) was not restricted to the progeny of heterozygote intercrossing, but was also observed in *Kit*^{tm1Alf/+} crosses with wild-type partners, independently of the gender combination (Table 1). It was not dependent on the genetic background of the mice, as the same phenotypes were found with the original 129/Sv *Kit*^{tm1Alf/+} heterozygotes and after at least six generations of backcrosses of the mutation onto the C57BL/6 and B6D2 (C57BL/6 $\times$ DBA/2) genetic backgrounds (Supplementary Table 1). The paramutated phenotype was inherited with a variable phenotypic extent depending on the crosses (Supplementary Fig. 1). It was most strongly expressed in second-generation crosses between *Kit*^{tm1Alf/+} heterozygotes, and still clearly recognizable in the progeny of *Kit*^{*} parents in the absence of the *tm1Alf* allele (*Kit*^{*} $\times$ *Kit*^{+/+} and *Kit*^{*} $\times$ *Kit*^{*} crosses; Table 1 and Supplementary Fig. 1). It would then progressively disappear in the following generations.

Another feature apparent from the data in Table 1 (see also Supplementary Fig. 1) is that in crosses between heterozygotes, the *Kit*^{tm1Alf/+} genotype was generated with frequencies in the range of 50–60% instead of the mendelian two-thirds. A probable explanation is that a fraction of the paramutable alleles have been modified in *Kit*^{tm1Alf/+} heterozygotes to the point of not being viable any more. The extent to which expression of the wild-type allele is altered varies between individuals, as indicated by the variable extent of the white fur patches (see Supplementary Fig. 1), and either the lack or an extensively reduced level of *Kit* receptor expression would not be compatible with normal development.

To probe the molecular basis of the paramutated phenotype, DNA and histone methylation were investigated in a CpG-rich region (nucleotides –31 to +219, with respect to the transcription start site)

¹Inserm, U636, F-06108 Nice, France. ²Université de Nice-Sophia Antipolis, Laboratoire de Génétique du Développement Normal et Pathologique, F-06108 Nice, France. ³Centre Commun de Microscopie Appliquée, Université de Nice-Sophia Antipolis, F-06108 Nice, France. [†]Present address: Unité de Génétique Moléculaire du Développement, Institut Pasteur, 75724 Paris Cedex, France.

that corresponds to the minimal *Kit* promoter⁶. Cytosine methylation was examined by amplification and sequencing after bisulphite treatment⁷. Possible changes in chromatin structure were investigated by chromatin immunoprecipitation with antibodies directed against the lysine 4 and lysine 9 dimethylated forms of histone H3, which are associated with active and repressed chromatin, respectively⁸. No significant change in either cytosine or histone methylation was observed between wild-type, heterozygous and paramutated animals (data not shown). However, we cannot exclude a critical role of differential DNA or histone methylation either in a specific cell type or in one of the more distant (and not yet precisely mapped) control regions that have been inferred from the analysis of *Kit* mutants⁹.

Reduced levels of polyadenylated RNA

The white-spotted phenotype of heterozygotes with a null mutant and a wild-type allele results from a reduced level of receptor expression⁵. We found that this was also the case in paramutated animals. Levels of polyadenylated *Kit* mRNA amounting to one-half of the wild-type homozygote were determined both in *Kit*^{tm1Alf/+} heterozygotes and in their paramutated (*Kit*^{*}) progeny (Fig. 2a), despite the presence in the latter of two structurally normal wild-type alleles. In addition to this marked decrease in mature mRNA, *Kit* RNA molecules of abnormal sizes accumulated; the possible origin of

these RNAs—abnormal arrest and/or initiation of primary transcription, abnormal post-transcriptional processing, and/or secondary cleavage of mature RNAs—remains to be determined. Starting in heterozygotes, distinct profiles of abnormal fragments were seen in different tissues. As shown in Fig. 2b, a prominent 0.37-kb RNA species was detected in the brain, and was identified (data not shown) as a spliced fragment of the mature *Kit* transcript including only exons 1 and 2. In the testis, northern analysis detected a more dispersed smear of RNA molecules of multiple sizes. As these abnormal short species were identified with a 5' probe corresponding to the region disrupted by *lacZ* insertion, and, on the other hand, did not hybridize with a *lacZ* probe, it is clear that they were derived from the genetically wild-type allele responsible for the *Kit*^{*} phenotype. They were clearly distinct from the transcript limited to the *lacZ* coding region expressed from the mutant allele (Fig. 2b).

RNA in *Kit*^{tm1Alf/+} sperm

The next question we addressed was the nature of the signal leading to the hereditary transfer of the paramutated state, first by a comparative analysis of spermatogenesis in the heterozygote and wild-type testis. A significant difference was noted in *Kit* transcription levels, with higher levels (determined by transcriptional run-on

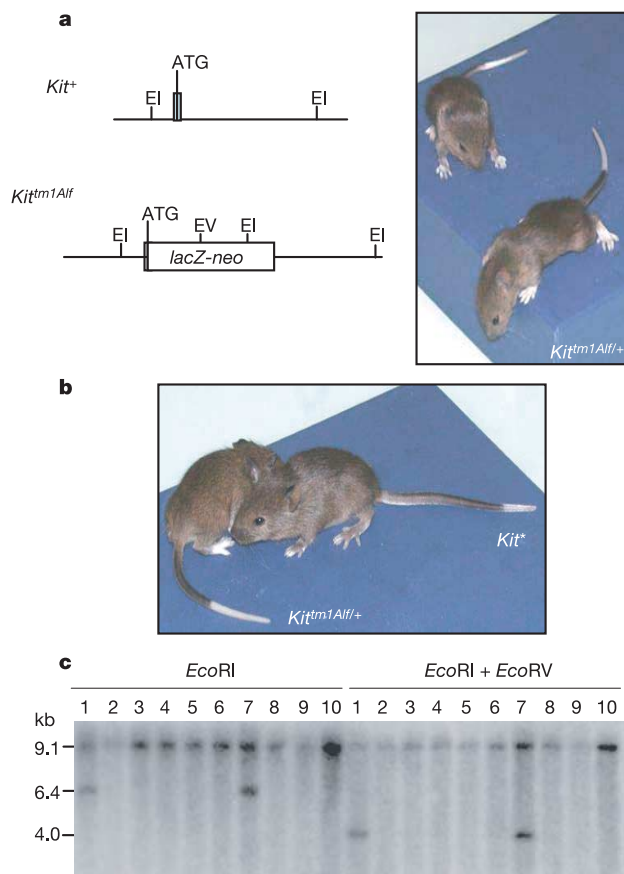


Figure 1 | The 'white-spotted' phenotype of *Kit*^{tm1Alf/+} heterozygotes and their paramutated progeny. **a**, Genotype and phenotype of the *Kit*^{tm1Alf/+} heterozygote. EI, *EcoRI* site; EV, *EcoRV* site. **b**, Heterozygote (*Kit*^{tm1Alf/+}) and paramutated (*Kit*^{*}) littermates. **c**, Nine mice with white tail tips from three litters were further analysed by Southern blot hybridization with a probe amplified from nucleotides -904 to +89 of the *Kit* locus⁵. Lanes 2–6, 8 and 9 show the wild-type genomic structure (*Kit*^{*} paramutated); lanes 1 and 7 show *Kit*^{tm1Alf/+} heterozygotes; lane 10 shows control B6D2 genomic DNA.

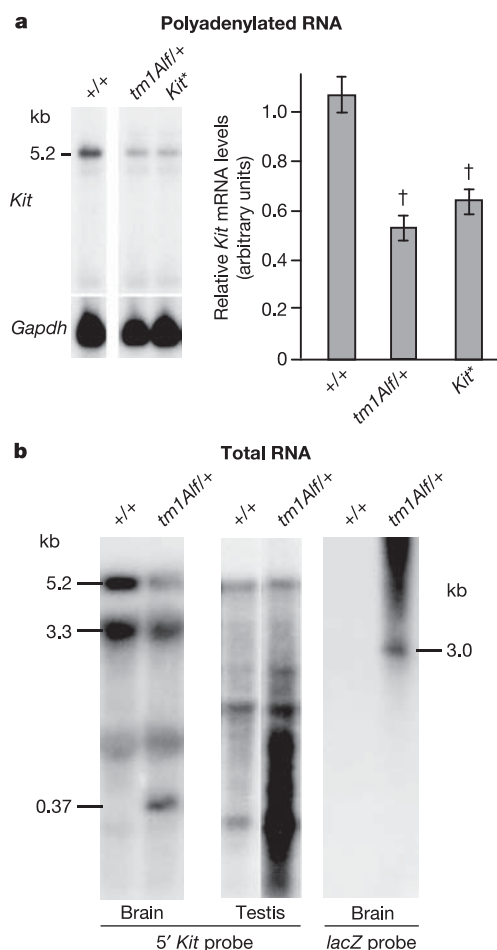


Figure 2 | Monoallelic levels of polyadenylated *Kit* RNA and abnormal patterns in total RNA. **a**, Northern analysis of brain mRNA from mice of the indicated genotype and densitometric measurements performed at successive exposure times of the *Kit* mRNA band relative to *Gapdh* mRNA in the same lane. Values are mean $\pm$ s.e.m.; †, significant difference ($P < 0.05$). **b**, Northern analysis of *Kit*^{tm1Alf/+} total RNA compared to *Kit*^{+/+} total RNA shows a variety of additional transcripts with different patterns in different organs. The 5' *Kit* probe covers exons 1 and 2, and the *lacZ* probe covers the entire coding region.

Table 1 | Segregation of coat phenotypes in the progeny of *Kit*^{tm1Alf/+} parents

Crosses†		Progeny‡			
Parents (male × female)	No. of crosses	Fur phenotype	<i>lacZ</i>	No. of mice	Class
<i>Kit</i> ^{tm1Alf/+} × <i>Kit</i> ^{tm1Alf/+}	8	White-spotted	+	30	<i>Kit</i> ^{tm1Alf/+}
		White-spotted	—	24	<i>Kit</i> [*]
		Full colour	—	3	<i>Kit</i> ^{+/+}
<i>Kit</i> ^{tm1Alf/+} × <i>Kit</i> ^{+/+}	4	White-spotted	+	16	<i>Kit</i> ^{tm1Alf/+}
		White-spotted	—	24	<i>Kit</i> [*]
		Full colour	—	4	<i>Kit</i> ^{+/+}
<i>Kit</i> ^{+/+} × <i>Kit</i> ^{tm1Alf/+}	4	White-spotted	+	22	<i>Kit</i> ^{tm1Alf/+}
		White-spotted	—	14	<i>Kit</i> [*]
		Full colour	—	5	<i>Kit</i> ^{+/+}
<i>Kit</i> [*] × <i>Kit</i> ^{+/+}	6	White-spotted (partially)§	ND	26	<i>Kit</i> [*]
		Full colour	ND	40	<i>Kit</i> ^{+/+}
		White-spotted (partially)§	ND	23	<i>Kit</i> [*]
<i>Kit</i> ^{+/+} × <i>Kit</i> [*]	5	Full colour	ND	34	<i>Kit</i> ^{+/+}

†B6D2 genetic background.

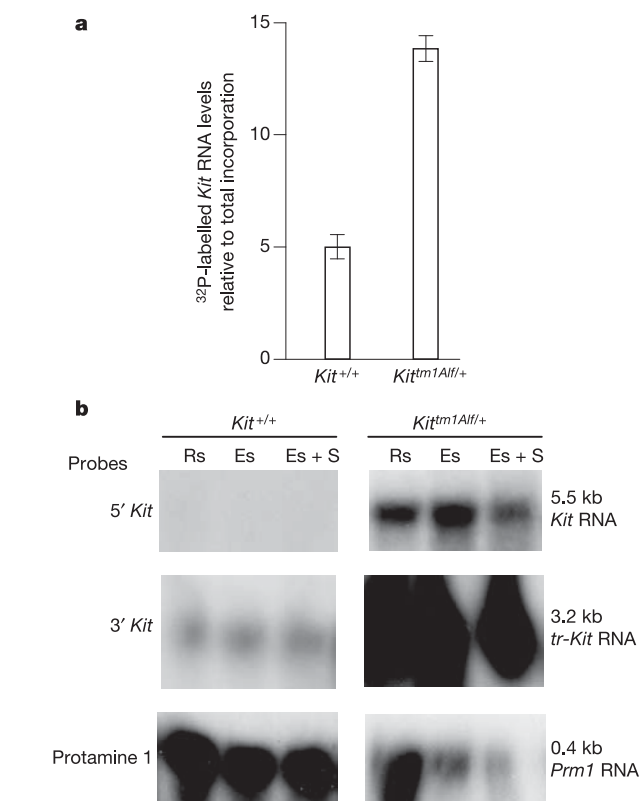
‡Phenotype: see Fig. 1b and Supplementary Fig. 1. Genotype: *lacZ* determination by genomic PCR amplification and expression by X-gal assay on tail. ND, not determined.§Reduced degree of white spotting (similar to *Kit*^{*} × *Kit*^{*} crosses; see Supplementary Fig. 1).

assays) found in heterozygotes (Fig. 3a). Deregulation was most obvious at the late spermatogenic stages. In wild-type mouse germ cells, *Kit* transcription is essentially restricted to spermatogonia, with reduced levels in early meiotic cells^{10–13}. The gene is virtually silent in the haploid phase, with the exception of a shorter RNA (*tr-Kit*) made

from an internal promoter in the most 3' region of the locus¹⁴. In contrast, northern blot analysis in *Kit*^{tm1Alf/+} germ cells showed significant amounts of 5' *Kit* RNA sequences in both round and elongated spermatids (Fig. 3b). This altered pattern of expression included an increased activity of both the upstream *Kit* and the internal *tr-Kit* promoters. The promoters of both alleles were affected, as high levels of β-galactosidase synthesis were evidenced in the haploid compartment of the tubules (Supplementary Fig. 2). These changes, clearly characteristic of the heterozygous state and at which paramutation is initiated, may be related to meiotic mispairing. Such an effect would be contrasting with the equally unexplained decrease in expression due to a perturbed synapsis that was described in a recent report¹⁵.

Most likely as a consequence of its deregulated expression, *Kit* RNA was detected not only in the spermatids of heterozygotes and paramutated mice, but, even more unexpectedly, in their mature epididymal sperm (Fig. 4). Both semiquantitative RT–PCR (PCR with reverse transcription) and quantitative real-time PCR on sperm extracts detected RNA sequences corresponding to the 5' region of the *Kit* gene. RT–PCR performed with only 20 cycles of amplification detected *Kit* RNA sequences (Fig. 4a), in addition to other transcripts (including *Gapdh*, *Prm1* and *Prm2*) (data not shown). These RNAs were never detected in the sperm of wild-type animals at such low cycle numbers. The presence of RNA molecules in human sperm has been reported previously (reviewed in ref. 16). However, the finding of increased amounts of *Kit* RNA in *Kit*^{tm1Alf/+} heterozygote sperm was not expected and needed confirmation. Acridine-orange staining showed two unexpected features in the sperm of heterozygous and paramutated males (Fig. 4b). Microscopic examination revealed the accumulation of yellow-stained material in the vicinity of the sperm nuclei (Fig. 4b, right panels), presumably corresponding to RNA¹⁷. Fluorescence-activated cell sorting (FACS; Fig. 4b, left panels) analysis confirmed a significant degree of yellow staining (vertical axis). However, a more variable intensity of green staining (corresponding to DNA; horizontal axis) was indicative of a less-compact chromatin structure.

RNA-containing structures can be identified by electron microscopy using the EDTA regressive staining technique, which is based on the chelation of uranyl ions by neutral EDTA (refs 18, 19). EDTA regressive staining in an Epon resin section showed densely contrasted structures corresponding to ribonucleoprotein constituents (Fig. 4c), whereas DNA-containing structures appeared greyish or bleached (see control spermatocyte sections in Fig. 4c). Enzymatic treatment of the sections (data not shown) verified that the EDTA regressive staining was abolished after RNase treatment and remained unchanged after extensive treatment of the sections with either DNase I or pronase. Spermatozoa in sections of

**Figure 3 | *Kit* RNA is overexpressed in heterozygote germ cells.**

a, Increased transcription rate. Incorporation of [α -³²P]UTP in permeabilized testicular cells from wild-type and heterozygote mice and radioactivity determination in hybridizations with a *Kit* probe, relative to total incorporation (see Methods). Error bars indicate s.e.m. of a total of ten assays on two animals of each genotype. **b**, Northern blot analysis with the 5' *Kit* probe of RNA prepared from elutriation fractions (90% pure¹³, starting from pools of 10 males of each genotype). Hybridization with a protamine 1 (*Prm1*) probe (control) and with the 3' *Kit* probe to detect the truncated form *tr-Kit*¹⁴ is also shown. Rs, round spermatids; Es, elongated spermatids; Es + S, a mixture of elongated spermatids and spermatozoa.

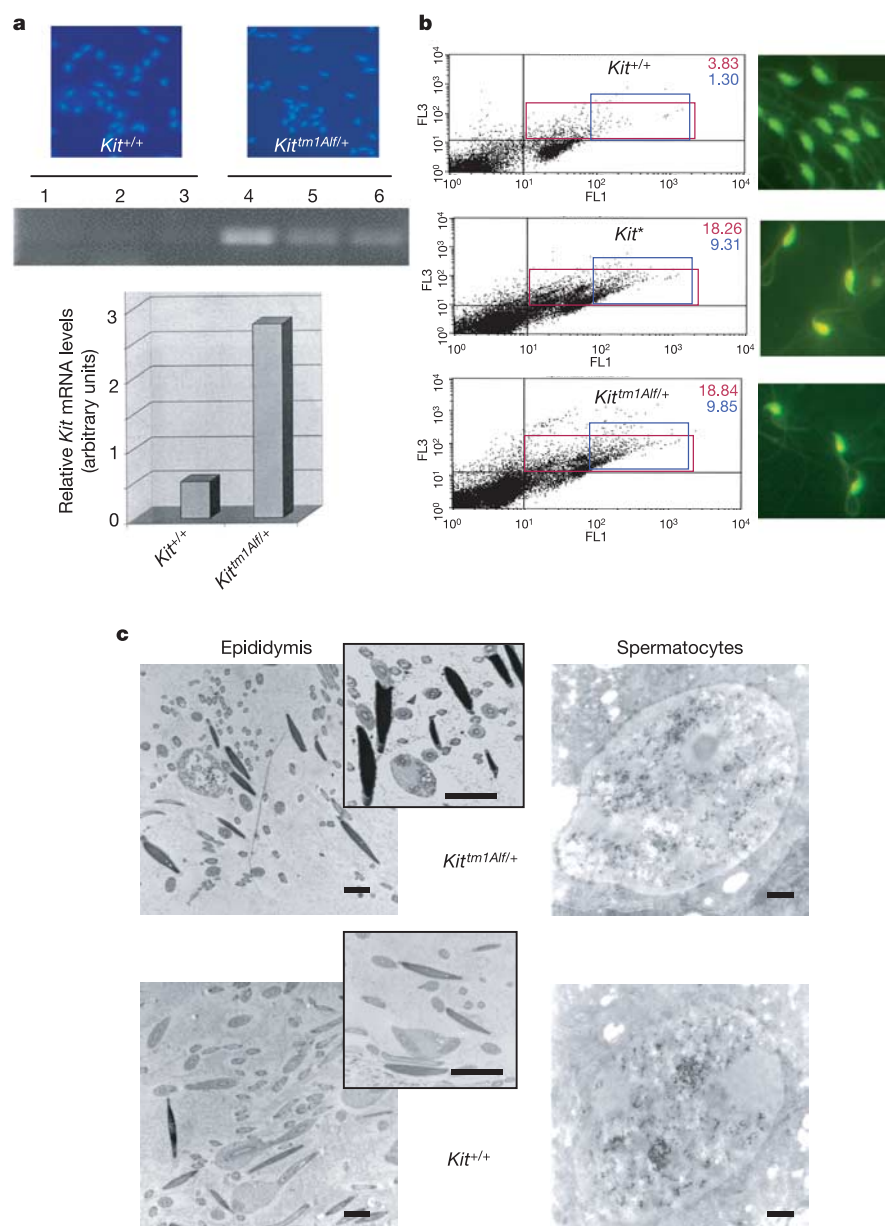


Figure 4 | RNA in *Kit*^{tm1Alf/+} spermatozoa. **a**, DAPI (4,6-diamidino-2-phenylindole) staining showing purity of the spermatozoa preparations, and RT-PCR determination of *Kit* RNA (lanes 1–3 show wild type; lanes 4–6 show heterozygotes) in spermatozoa. Quantitative real-time PCR determination of *Kit* RNA relative to *Gapdh* RNA. **b**, Acridine-orange staining followed by fluorescence microscopy (right panels) and FACS analysis (left panels) of RNA (yellow, vertical axis) and DNA (green, horizontal axis). Values in red and blue are the proportions of cells in the corresponding compartments. **c**, EDTA regressive staining. Top panels are *Kit*^{tm1Alf/+}; bottom panels are *Kit*^{+/+}; left panels are epididymis (insert shows higher magnification); right panels are control staining of spermatocytes showing the greyish appearance of nuclei (DNA) and the contrasted stain (RNA) in the cytoplasm. Scale bars, 1 μ m.

epididymis of *Kit*^{tm1Alf/+} heterozygotes showed more heterogeneous shapes and more contrasted staining than in wild-type males (Fig. 4c). The generally light staining of the wild-type mouse sperm by the EDTA regressive staining technique, with only the rare occurrence of somewhat more-contrasted sections (Fig. 4c), may reflect a physiological low level of RNA, as reported for human sperm¹⁶. Taken together with the RT-PCR results and with the deregulated expression of *Kit* at the late spermatogenic stages, these observations are indicative of the presence of unusual amounts of RNA.

RNA induction

The presence of RNA in sperm cells led us to consider the possibility that transfer of RNA to the fertilized egg could be the signal leading to the paramutated phenotype. Although the molecular mechanisms involved would remain to be established, such a hypothesis would be consistent with recent examples of RNA molecules responsible for stable epigenetic changes (reviewed in ref. 20). A series of experiments were performed in which RNA prepared from either *Kit*^{+/+} homozygotes or *Kit*^{tm1Alf/+} heterozygotes was microinjected into B6D2 one-cell embryos following standard procedures for DNA

microinjection²¹. No toxicity was noted and, in every litter born after *Kit*^{tm1Alf/+} RNA injection, close to 50% of the offspring showed the white tail tip characteristic of the heterozygote (Fig. 5). The same result was achieved after injection of either somatic (brain) RNA or RNA prepared from heterozygotic sperm. We noted, however, the occurrence of rare white-spotted mice in the control litters produced after microinjection of RNA prepared from wild-type brain, as well as after injection of irrelevant (*lacZ*) RNA. Nevertheless, we concluded that the white-spotted phenotype observed after microinjection of *Kit*^{tm1Alf/+} RNA was a specific effect of the heterozygote RNA for two reasons. First, its frequency in control groups was significantly lower than with heterozygote RNA, with smaller areas of white fur. Second, and more significantly, the rare white tail phenotype of the controls was either very inefficiently, or not at all, transmitted to the progeny in crosses with wild-type partners, in clear contrast with the phenotype induced by *Kit*^{tm1Alf/+} RNA, which was efficiently transmitted to subsequent generations (Fig. 5 and Supplementary Table 3).

Taken together with the presence of RNA molecules of abnormal sizes in preparations from heterozygotes (Fig. 2), we proposed that the paramutated state was induced by a partial degradation product of *Kit* RNA. We attempted to induce RNA degradation by injecting

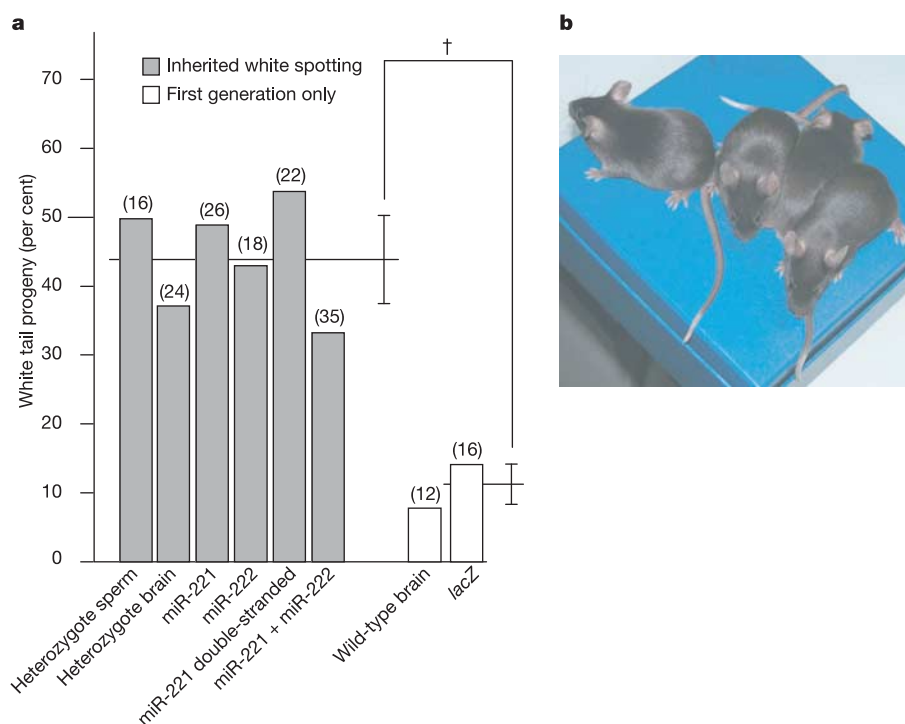


Figure 5 | A heritable mutant-like phenotype induced by RNA microinjection in one-cell embryos. **a**, White-spotted progeny (per cent; total number is shown in parentheses at the top of each bar) born after microinjection of the indicated RNAs. Sequences of oligonucleotide primers and miRNAs are listed in Supplementary Information. Horizontal lines show the average $\pm$ s.e.m. of each group ($\dagger$, $P < 0.05$); open bars indicate a low efficiency of transmission ($\leq 3\%$); closed bars indicate efficiencies of transmission of 56–78% (complete data are listed in Supplementary Table 2). **b**, F₁ progeny of a wild-type female mated with a white-spotted male born after injection of *Kit*^{tm1Alfj} brain RNA.

either one of the two microRNAs (miRNAs), miR-221 and miR-222, that had been identified as potentially targeting *Kit* mRNA in a computational survey of mammalian microRNAs²², to test whether injection of these two miRNAs could have the same effect. This experiment showed that this was indeed the case, with the white tail phenotype being induced at high frequency and being efficiently inherited (Fig. 5b; Supplementary Table 2). Exposure to microRNAs of the early embryonic genome thus seems to be sufficient to induce a permanent and heritable epigenetic change in gene expression.

Conclusion

Although it is tempting to think in terms of causal relationships, the mechanisms leading to the inherited modification of the paramutated phenotype, and to the similar phenotypes induced by exposure to the abnormal species of *Kit* RNA and to microRNAs, remain to be determined. Further characterization of the RNAs in heterozygotes, their effects when injected into zygotes, and the mechanistic aspects of the chromatin remodelling processes directed by microRNAs will hopefully lead to a more complete and better-defined picture.

The initial event inducing paramutation is not known, even in the most thoroughly investigated plant systems². Incomplete meiotic pairing of homologous chromosomes is considered as the determining event in the epigenetic changes known as co-suppression in plants and meiotic silencing in *Neurospora*²³. Preliminary results on two other *Kit* mutants would be consistent with this view. We generated a distinct insertion mutant carrying a green fluorescent protein (*GFP*)-*neo* cassette in the first intron of the *Kit* gene and found that paramutated progeny were generated under the same conditions and with the same frequency as in *Kit*^{tm1Alfj} crosses. In contrast, strictly mendelian phenotypic and genotypic segregations were found consistently in the offspring of the classical point-mutant *Kit*^{W-v} (ref. 24; data not shown).

A number of epigenetic determinations are currently under study in various systems, and a role for RNA has been suggested in several instances (reviewed in ref. 2): the induction of heritable phenotypic changes by double-stranded RNA was reported in *Caenorhabditis elegans*²⁵, and an RNA cache was suggested as a carrier of genetic information in *Arabidopsis*²⁶. On the other hand, one of

the challenging aspects of the current results is the hereditary transmission of epigenetic states. Paternal and maternal transmission were equally efficient. Analyses of the sperm cell—a simpler structure than the oocyte—led us to conclude that it is not only a vector for the male haploid genome, but also the carrier of additional information in the form of RNA molecules. The EDTA regressive staining technique, which allows the analysis of spermatozoa at the cellular level, and the microinjection of RNA and microRNAs into fertilized eggs should be useful tools to enable a functional analysis. The mouse model might then provide a clue as to the function of the RNA molecules observed in human sperm¹⁶. The hypothesis that RNAs of paternal origin, including microRNAs, can have a role in modulating gene expression in the embryo has been recently formulated (reviewed in ref. 16), and paramutation in the *Kit* gene may provide a useful experimental model for further analysis.

METHODS

Mice and genotyping. *Kit*^{tm1Alfj} heterozygotes were initially obtained from J. J. Panthier (Institut Pasteur, Paris). Three stocks were generated and maintained in parallel—one on the original 129/Sv genetic background, and two by crosses onto C57BL/6 and C57BL/6 $\times$ DBA/2 (B6D2) backgrounds (in both cases for at least six generations). We determined genotypes by PCR assays specific for the *neo* and *lacZ* transgenes and by Southern blot hybridization with a genomic probe. Investigations were conducted in accordance with French and European regulations for the care and use of research animals.

Southern blot analysis. Analysis was performed after cleavage with *EcoRI* and *EcoRV* enzymes as described⁵.

RNA analysis. Northern analysis was performed as previously described³. Polyadenylated RNA was prepared from total RNA using the mRNA isolation kit (Roche) according to the manufacturer's instructions. The 5' probe for the detection of *Kit* mRNA covered the distal part of exon 1 and exon 2, from nucleotides 69 to 374 of the complementary DNA sequence (GenBank accession number AY536430). The probe for the 3' part of the *Kit* transcript and the truncated *tr-Kit* transcript¹⁴ extended from nucleotides 2418 to 2776 (see Supplementary Information for the nucleotide sequences). Quantification was performed by densitometric analysis of autoradiographs at various exposure times. *Kit* RNA levels were normalized to the level of *Gapdh* mRNA. Quantitative PCR assays were performed with the ABI Prism apparatus (Applied Biosystems) and the Syber Green I kit (Eurogentec). Sequences of oligonucleotide primers are provided in Supplementary Information.

Transcriptional run-on assays. Assays of transcriptional activity by

radiolabelling of RNA transcripts with [α - 32 P]UTP in isolated nuclei²⁷ were performed on testicular cell preparations from 10-week-old males. Five independent assays per animal were performed on two animals of each genotype. A detailed procedure is provided in Supplementary Information.

Acridine-orange staining and FACS analysis. Preparation of spermatozoa from the epididymis of wild-type, heterozygote and paramutated males were fixed in buffered 10% formaldehyde for 30 min, rinsed with phosphate-buffered saline, treated with 0.2% trypsin and 0.01% Triton X-100 for 30 min at room temperature (20 °C), washed and stained with acridine orange (Sigma) for 15 min according to published procedures¹⁷. Samples were analysed on a CAS200 cell analysis system (Becton-Dickinson).

Microinjection into fertilized eggs. RNA microinjection into B6D2 fertilized eggs (after normal ovulation) was performed by the standard techniques of DNA injection²¹. One to two picolitres of a 10 μ g ml⁻¹ solution of total RNA or 0.1 μ g ml⁻¹ solutions of RNA oligonucleotides in 5 mM Tris-HCl, pH 8.0, 0.1 mM EDTA were injected.

In situ determination of β -galactosidase activity. X-gal staining of β -galactosidase activity was performed as previously described²⁸.

Electron microscopy. Mouse testes and epididymes were fixed immediately after dissection in 1.6% glutaraldehyde in 0.1 M phosphate buffer (1 h at 4 °C). They were rinsed with buffer and free aldehyde groups were blocked with 50 mM NH₄Cl in PBS for 30 min at 4 °C. Specimens were dehydrated with acetone and embedded in Epon resin. RNA-protein complexes were visualized by the EDTA regressive staining technique¹⁸. Shortly after, grids were stained for 1 min with 4% aqueous solution of uranyl acetate (at 18 °C) and treated for 30 min in 0.2 M EDTA, pH 7.0. Grids were carefully rinsed with distilled water and stained for 1 min with lead citrate. Under these conditions, only RNA molecules remained stained. All grids were observed in a Philips CM12 electron microscope operating at 60 or 80 kV and equipped with a 30- μ m objective aperture. All recording films were taken and treated under similar working conditions.

Received 24 September 2005; accepted 20 February 2006.

- Brink, R. A. A genetic change associated with the *R* locus in maize which is directed and potentially reversible. *Genetics* **41**, 872–879 (1956).
- Chandler, V. L. & Stam, M. Chromatin conversations: mechanisms and implications of paramutation. *Nature Rev. Genet.* **5**, 532–544 (2004).
- Rassoulzadegan, M., Magliano, M. & Cuzin, F. Transvection effects involving DNA methylation during meiosis in the mouse. *EMBO J.* **21**, 4404–4450 (2002).
- Herman, H. *et al.* Trans allele methylation and paramutation-like effects in mice. *Nature Genet.* **34**, 199–202 (2003).
- Bernex, F. *et al.* Spatial and temporal patterns of *c-kit*-expressing cells in *W^{lacZ/+}* and *W^{lacZ}/W^{lacZ}* mouse embryos. *Development* **122**, 3023–3033 (1996).
- Yasuda, H., Galli, S. J. & Geissler, E. N. Cloning and functional analysis of the mouse *c-kit* promoter. *Biochem. Biophys. Res. Commun.* **191**, 893–901 (1993).
- Olek, A., Oswald, J. & Walter, J. A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* **24**, 5064–5066 (1996).
- Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14**, 286–298 (2002).
- Kluppel, M., Nagle, D. L., Bucan, M. & Bernstein, A. Long-range genomic rearrangements upstream of *Kit* dysregulate the developmental pattern of *Kit* expression in *W⁵⁷* and *W^{banded}* mice and interfere with distinct steps in melanocyte development. *Development* **124**, 65–77 (1997).
- Manova, K., Nocka, K., Besmer, P. & Bachvarova, R. F. Gonadal expression of *c-kit* encoded at the *W* locus of the mouse. *Development* **110**, 1057–1069 (1990).
- Motro, B., van der Kooy, D., Rossant, J., Reith, A. & Bernstein, A. Contiguous patterns of *c-kit* and *steel* expression: analysis of mutations at the *W* and *Sl* loci. *Development* **113**, 1207–1221 (1991).
- Sorrentino, V., Giorgi, M., Geremia, R., Besmer, P. & Rossi, P. Expression of the *c-kit* proto-oncogene in the murine male germ cells. *Oncogene* **6**, 149–151 (1991).
- Vincent, S. *et al.* Stage-specific expression of the *Kit* receptor and its ligand (KL) during male gametogenesis in the mouse: a *Kit*-*KL* interaction critical for meiosis. *Development* **125**, 4585–4593 (1998).
- Albanesi, C. *et al.* A cell- and developmental stage-specific promoter drives the expression of a truncated *c-kit* protein during mouse spermatid elongation. *Development* **122**, 1291–1302 (1996).
- Turner, J. M. *et al.* Silencing of unsynapsed meiotic chromosomes in the mouse. *Nature Genet.* **37**, 41–47 (2005).
- Krawetz, S. A. Paternal contribution: new insights and future challenges. *Nature Rev. Genet.* **6**, 633–642 (2005).
- Darzynkiewicz, Z. Differential staining of DNA and RNA in intact cells and isolated cell nuclei with acridine orange. *Methods Cell Biol.* **33**, 285–298 (1990).
- Bernhard, W. A new staining procedure for electron microscopical cytology. *J. Ultrastruct. Res.* **27**, 250–265 (1969).
- Biggio, M. & Fakan, S. Fine structural specific visualization of RNA on ultrathin sections. *J. Histochem. Cytochem.* **46**, 389–395 (1998).
- Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev.* **19**, 1635–1655 (2005).
- Hogan, B., Beddington, R., Costantini, F. & Lacy, L. *Manipulating the Mouse Embryo: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1994).
- Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
- Shiu, P. K., Raju, N. B., Zickler, D. & Metzner, R. L. Meiotic silencing by unpaired DNA. *Cell* **107**, 905–916 (2001).
- Nocka, K. *et al.* Molecular bases of dominant negative and loss of function mutations at the murine *c-kit*/white spotting locus: *W³⁷*, *W^v*, *W⁴¹* and *W*. *EMBO J.* **9**, 1805–1813 (1990).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Lolle, S. J., Victor, J. L., Young, J. M. & Pruitt, R. E. Genome-wide non-mendelian inheritance of extra-genomic information in *Arabidopsis*. *Nature* **434**, 505–509 (2005).
- Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2001).
- Sage, J. *et al.* Temporal and spatial control of the *Sycp1* gene transcription in the mouse meiosis: regulatory elements active in the male are not sufficient for expression in the female gonad. *Mech. Dev.* **80**, 29–39 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are indebted to K. Thyagarajan for her participation in experimental work, J. J. Panthier for the gift of *Kit^{tm1Alf/+}* mice and for helpful discussions, and K. B. Marcu and A. Schedl for help in preparing the manuscript. We thank M. Aupetit, Y. Fantei-Caujolle, J. P. Laugier, S. Pagnotta and K. Rassoulzadegan for expert technical assistance. This work was made possible by a grant to M.R. as 'Equipe Labellisée' of the 'Ligue Nationale Française Contre le Cancer'.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.R. (minoo@unice.fr).

Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells

Ömer H. Yilmaz¹, Riccardo Valdez², Brian K. Theisen², Wei Guo³, David O. Ferguson², Hong Wu³ & Sean J. Morrison¹

Recent advances have highlighted extensive phenotypic and functional similarities between normal stem cells and cancer stem cells. This raises the question of whether disease therapies can be developed that eliminate cancer stem cells without eliminating normal stem cells. Here we address this issue by conditionally deleting the *Pten* tumour suppressor gene in adult haematopoietic cells. This led to myeloproliferative disease within days and transplantable leukaemias within weeks. *Pten* deletion also promoted haematopoietic stem cell (HSC) proliferation. However, this led to HSC depletion via a cell-autonomous mechanism, preventing these cells from stably reconstituting irradiated mice. In contrast to leukaemia-initiating cells, HSCs were therefore unable to maintain themselves without *Pten*. These effects were mostly mediated by mTOR as they were inhibited by rapamycin. Rapamycin not only depleted leukaemia-initiating cells but also restored normal HSC function. Mechanistic differences between normal stem cells and cancer stem cells can thus be targeted to deplete cancer stem cells without damaging normal stem cells.

Cancer stem cells have notable phenotypic and mechanistic similarities to normal stem cells in the same tissues^{1–4}. Acute myeloid leukaemia (AML) is sustained by leukaemic stem cells that are also called leukaemia-initiating cells because they are defined by their ability to transfer disease on transplantation into irradiated mice^{5–7}. Leukaemia-initiating cells express markers similar to normal HSCs^{5,6} and depend on similar mechanisms to self-renew^{8,9}. Brain cancer stem cells also express markers of normal neural stem cells and depend on similar pathways for their proliferation^{4,10}. The Hedgehog, Wnt and Notch pathways that often promote cancer cell proliferation also promote normal stem cell self-renewal^{1,2,11,12}. Conversely, tumour suppressors that inhibit cancer cell proliferation—such as p53, p16^{INK4a} and p19^{ARF}—also inhibit stem cell self-renewal^{11,13,14}. Whether cancer stem cells arise from normal stem cells or other cells, their similarity to normal stem cells indicates that they inherit or acquire stem cell properties. This raises the question of whether it will be possible to identify therapies that eliminate cancer stem cells without eliminating normal stem cells in the same tissues.

We have addressed this issue by examining the effect of *Pten* deletion on leukaemia-initiating cells and normal HSCs. PTEN (for phosphatase and tensin homologue) is a phosphatase that negatively regulates signalling through the phosphatidylinositol-3-OH kinase (PI(3)K) pathway, inhibiting proliferation and survival^{15,16}. *Pten* is commonly deleted or otherwise inactivated in diverse cancers¹⁷, including haematopoietic malignancies^{18–21}. Here we report that whereas *Pten* deletion causes the generation of transplantable leukaemia-initiating cells, it also causes the depletion of normal HSCs, thus identifying a mechanistic difference between the maintenance of normal stem cells and cancer stem cells.

Pten deletion leads to leukaemogenesis

Pten was conditionally deleted from 6-to-8-week-old *Pten*^{fl/fl}; *Mx-1-Cre* mice by administering seven doses of polyinosine-polycytidine (pIpC) over 14 days to induce Cre expression^{22,23}. After 14 days, *Pten* seemed to be completely deleted from HSCs and other haematopoietic cells (Supplementary Fig. 1). We analysed *Pten*^{fl/fl}; *Mx-1-Cre* mice, as well as *Pten*^{fl/+}; *Mx-1-Cre* littermate control mice, five days after pIpC treatment. Almost all *Pten*^{fl/fl}; *Mx-1-Cre* mice (17 out of 19) developed myeloproliferative disease marked by a tenfold increase in spleen cellularity (Fig. 1c), complete histological effacement of the splenic architecture (Fig. 1b), reduced bone marrow cellularity (Fig. 1c), and increased blast cell frequency (Fig. 1d). The increased spleen cellularity was largely attributable to extramedullary haematopoiesis (Supplementary Fig. 2c, d) with a prominent expansion in the number of immature myeloid cells (Supplementary Fig. 2e–g; Supplementary Table 1). None out of 20 *Pten*^{fl/+}; *Mx-1-Cre* littermates showed these changes after pIpC treatment (Fig. 1c, a, d).

Within 4 to 6 weeks after pIpC treatment, most *Pten*^{fl/fl}; *Mx-1-Cre* mice progressed to frank leukaemia²⁴, including AML and acute lymphoblastic leukaemia (ALL), and died (for the criteria used to diagnose leukaemias, see Supplementary Table 2). AMLs were characterized by large numbers of chloroacetate-esterase-positive myeloid blasts in the spleen (Fig. 1e), and ALLs were characterized by large numbers of terminal deoxynucleotidyl transferase (TdT)-positive lymphoid blasts throughout the thymus, which was also enlarged and effaced (Fig. 1f). The bone marrow contained Mac-1⁺Gr-1^{low}CD4[−] myeloid blasts and CD4⁺CD8⁺CD3⁺Mac-1[−] lymphoid blasts (Fig. 1h, i; data not shown). Karyotypic analysis of myeloid blasts from four *Pten*-deleted mice with AML revealed

¹Howard Hughes Medical Institute, Life Sciences Institute, Department of Internal Medicine, and Center for Stem Cell Biology, University of Michigan, Ann Arbor, Michigan 48109-2216, USA. ²Department of Pathology, University of Michigan, Ann Arbor, Michigan 48109-2216, USA. ³Molecular and Medical Pharmacology, UCLA School of Medicine, Los Angeles, California 90095-1735, USA.

marked aneuploidy and/or chromosomal translocations, suggesting leukaemogenesis was associated with additional mutations after *Pten* deletion (Supplementary Fig. 3). Common abnormalities across multiple karyotypes from each mouse suggested that the AMLs were clonal or oligoclonal (Supplementary Fig. 3).

HSCs proliferate after *Pten* deletion but become depleted

We examined the cell cycle status of whole bone marrow cells and Flk-2⁺Sca-1⁺Lin⁺c-Kit⁺CD48⁺ cells five days after completing pIpC administration. Flk-2⁺Sca-1⁺Lin⁺c-Kit⁺CD48⁺ cells are highly enriched for HSCs^{25–27}, representing only 0.005% of bone marrow cells. Nineteen out of 20 mice transplanted with 15 Flk-2⁺Sca-1⁺Lin⁺c-Kit⁺CD48⁺ cells from control mice showed long-

term multilineage reconstitution by the donor cells (Supplementary Table 3). There was no significant effect of *Pten* deletion on the cell cycle distribution of whole bone marrow cells (Fig. 2a, b). In contrast, there was a three-to-four-fold increase in the percentage of dividing Flk-2⁺Sca-1⁺Lin⁺c-Kit⁺CD48⁺ cells, and in the percentage that incorporated the nucleotide analogue BrdU (5-bromodeoxyuridine) over a 19-h period (Fig. 2c–f). These data suggest that *Pten* promotes quiescence in HSCs, and that in the absence of *Pten* HSCs are driven into cycle.

Consistent with this, the absolute number of Flk-2⁺Sca-1⁺Lin⁺c-Kit⁺CD48⁺ cells per *Pten*^{fl/fl}; *Mx-1-Cre* mouse increased by approximately threefold within five days of pIpC treatment (Fig. 2g; see Methods for details). Notably, when *Pten*^{fl/fl}; *Mx-1-Cre* mice were

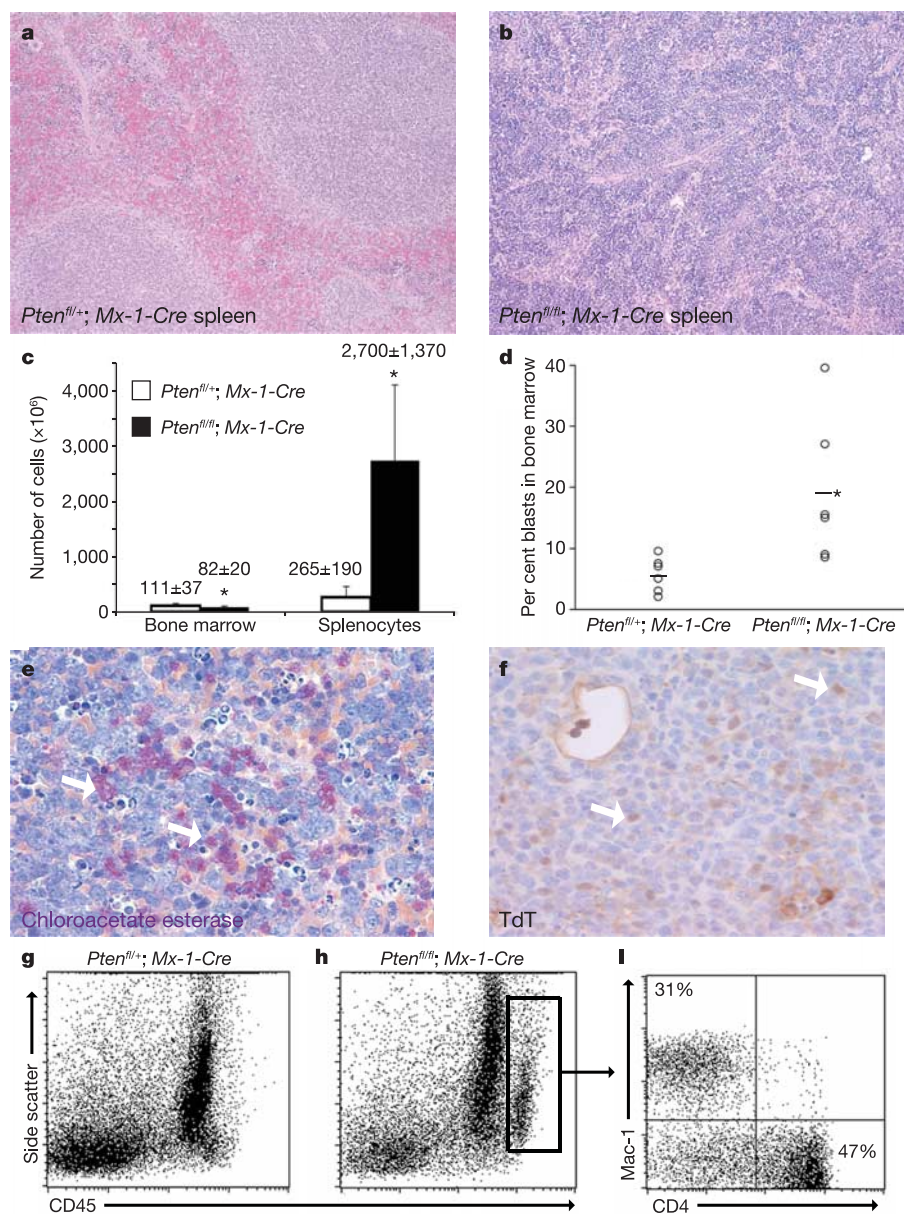


Figure 1 | *Pten* deletion from adult haematopoietic cells leads to myeloproliferative disease that progresses to AML and ALL. **a–c**, Within five days of *Pten* deletion, *Pten*^{fl/fl}; *Mx-1-Cre* mice developed myeloproliferative disease marked by increased spleen cellularity (**c**), reduced bone marrow cellularity (**c**), and complete effacement of the splenic architecture (compare **b** with **a**) owing to myeloid-predominant extramedullary haematopoiesis (Supplementary Fig. 2). Values are mean $\pm$ s.d. *, $P < 0.05$. **d**, Blast cell frequency was significantly increased in *Pten*^{fl/fl}; *Mx-1-Cre* mice, but five days after pIpC treatment only a minority

of these mice (2 out of 6) showed more than 20% blasts. Bars indicate mean values. *, $P < 0.01$. **e, f**, By six weeks after pIpC treatment, most *Pten*^{fl/fl}; *Mx-1-Cre* mice had greater than 20% blast cells in the bone marrow, were acutely ill, and showed the pathological features of both AML (**e**; mononuclear blasts in the spleen; white arrows) and ALL (**f**; TdT-positive lymphoid blasts in the thymus; white arrows). **g–i**, The bone marrow of *Pten*^{fl/fl}; *Mx-1-Cre* mice contained a blast cell population (box in **h**), with myeloid (**i**; Mac-1-positive) and lymphoid (**i**; CD4-positive) cells, not evident in control mice (**g**).

examined 24 to 39 days after pIpC treatment, the number of Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ cells declined significantly below that found in control mice (Fig. 2g). This suggested that HSCs transiently expanded in number after *Pten* deletion, but were unable to maintain themselves and subsequently became depleted.

To test the function of *Pten*-deficient HSCs, we transplanted 15 Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ donor cells from *Pten*^{fl/fl}; *Mx-1-Cre* mice or littermate controls (five days after pIpC) into irradiated recipient mice along with 200,000 recipient bone marrow cells (Fig. 2h). Whereas control cells gave high levels of long-term multilineage reconstitution in all five recipients, *Pten*-deficient cells initially gave multilineage reconstitution, but by eight weeks after transplantation none of the seven recipients remained multilineage-reconstituted. None of these recipients developed AML or ALL. Small numbers of *Pten*-deficient HSCs were thus capable of efficiently engrafting and undergoing multilineage differentiation but became depleted over time and gave only transient, rather than long-term, multilineage reconstitution. In four independent experiments, a total of 23 mice that were injected with 15 Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ cells from *Pten*^{fl/fl}; *Mx-1-Cre* donors did not develop leukaemia. Eighteen of these mice showed transient multilineage reconstitution whereas only two remained long-term multilineage-reconstituted, and levels of reconstitution declined continuously in

both of these mice (Supplementary Table 3). The fact that donor chimaerism consistently declined over time, even in mice that did not develop leukaemia, demonstrates that the depletion of *Pten*-deficient HSCs is not caused by leukaemogenesis.

We performed similar experiments with whole bone marrow cells. Whereas all recipients of control bone marrow cells showed long-term multilineage reconstitution (Fig. 2i) with high levels of donor cells (Fig. 2j), the percentage of recipients that were multilineage-reconstituted by *Pten*^{fl/fl}; *Mx-1-Cre* cells, and the levels of reconstitution in these recipients, declined over time. The results with whole bone marrow rule out the possibility that *Pten*-deficient HSCs simply changed their surface-marker phenotype. Moreover, only 3 out of 10 recipients of *Pten*^{fl/fl}; *Mx-1-Cre* bone marrow cells developed AML or ALL, further demonstrating that the loss of HSC activity was not secondary to leukaemogenesis.

Pten acts cell-autonomously to maintain HSCs

The possibility remained that a few days of exposure to the myeloproliferative disease in donor mice might have irreversibly damaged the *Pten*-deficient HSCs before transplantation. To test whether the depletion of HSCs in *Pten*^{fl/fl}; *Mx-1-Cre* mice was an indirect consequence of neoplasms/altered haematopoietic environment, or whether *Pten* is required cell-autonomously for the maintenance of

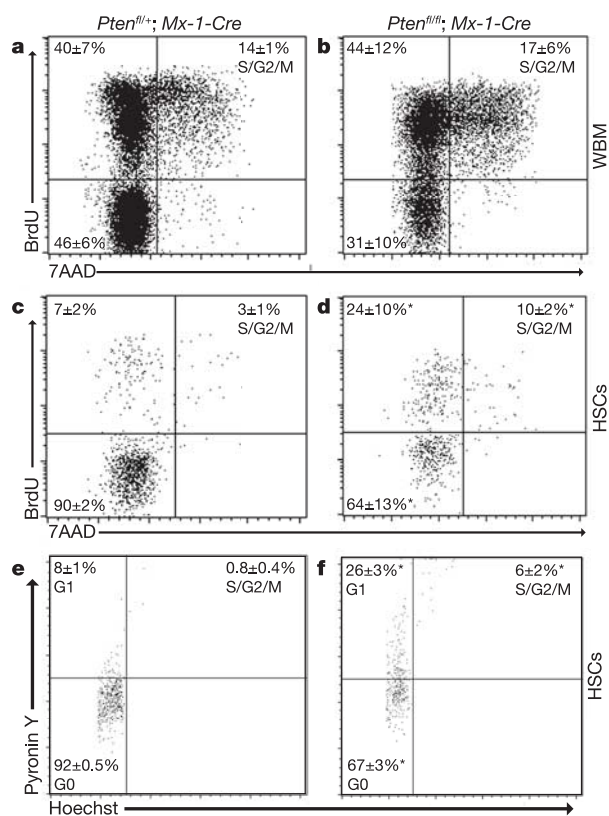


Figure 2 | HSCs proliferate after *Pten* deletion, transiently expanding in number before becoming depleted. **a, b**, No significant difference in the cell cycle status of whole bone marrow (WBM) cells from *Pten*^{fl/fl}; *Mx-1-Cre* mice (**b**) and littermate controls (**a**) was observed five days after pIpC treatment. BrdU was administered for 19 h to mark cells that entered S phase, and 7-amino-actinomycin D (7-AAD) indicated DNA content. Values are mean ± s.d. of three experiments. **c, d**, Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ HSCs from *Pten*^{fl/fl}; *Mx-1-Cre* mice (**d**) included significantly (*, $P < 0.05$) more dividing cells relative to controls (**c**). Values are mean ± s.d. of three experiments. **e, f**, Significantly (*, $P < 0.01$) fewer Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ cells from *Pten*^{fl/fl}; *Mx-1-Cre* mice (**f**) were in G0 compared with controls (**e**) on the basis of Hoechst/Pyrin Y staining of DNA/RNA content. Values are mean ± s.d. of three experiments. **g**, Total number of Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ cells in *Pten*^{fl/fl}; *Mx-1-Cre* mice

was significantly increased relative to littermate controls five days after pIpC treatment, but significantly decreased relative to controls 24 to 39 days after pIpC administration. Values are mean ± s.d. of six experiments. **h**, Fifteen donor Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ cells from *Pten*^{fl/fl}; *Mx-1-Cre* mice or controls were transplanted into irradiated recipients along with 200,000 recipient bone marrow cells. Whereas control cells gave long-term multilineage reconstitution in all recipients ($n = 5$), *Pten*-deficient cells gave only transient multilineage reconstitution ($n = 7$; *, $P < 0.05$). **i, j**, Similar results were observed when 300,000 bone marrow cells from *Pten*^{fl/fl}; *Mx-1-Cre* mice or controls were transplanted into irradiated recipients ($n = 10$ and 4, respectively; *, $P < 0.05$) along with 300,000 recipient bone marrow cells. In **h** and **j**, error bars represent standard deviation.

HSCs, we transplanted *Pten*^{fl/fl}; *Mx-1-Cre* bone marrow cells or control bone marrow cells (both CD45.2⁺) into recipient mice (CD45.1⁺) along with half as many recipient bone marrow cells (Fig. 3a). Six weeks after transplantation, when these mice showed stable chimaerism, *Pten* was deleted and the relative frequencies of donor and recipient HSCs were monitored over time. As expected, two days after pIpC treatment donor cells accounted for roughly two-thirds of HSCs in the bone marrow irrespective of whether they were from control (Fig. 3b) or *Pten*-deficient (Fig. 3c) mice.

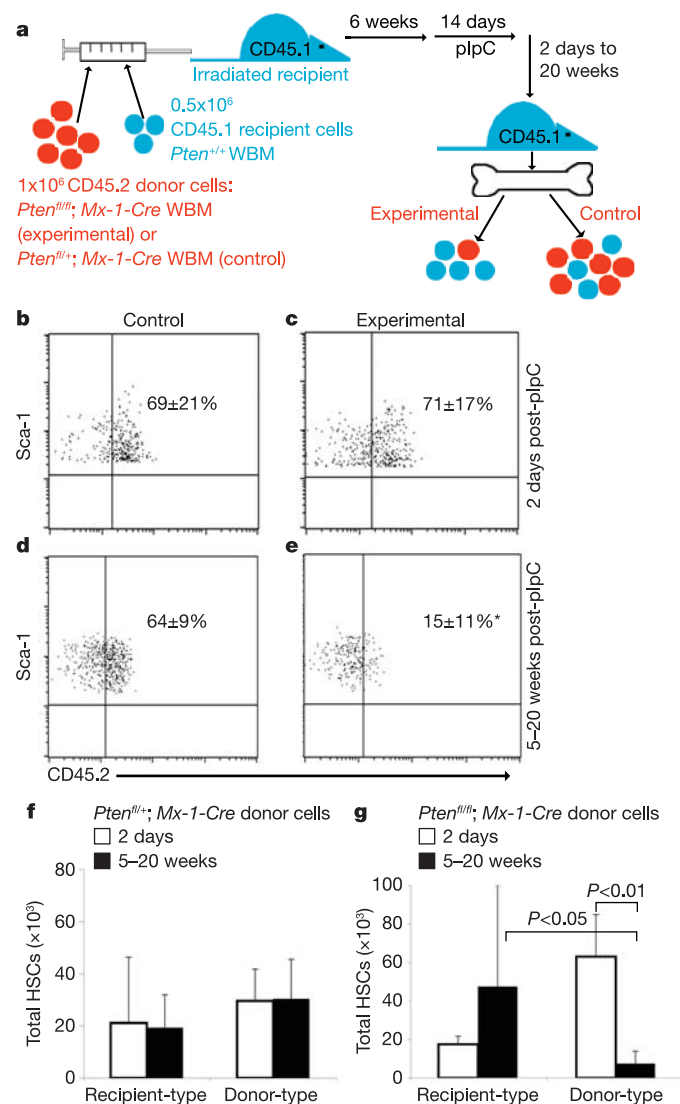


Figure 3 | *Pten* is required cell-autonomously for HSC maintenance.

a, Irradiated (CD45.1) recipients were transplanted with a 2:1 ratio of donor (CD45.2) *Pten*^{fl/fl}; *Mx-1-Cre* bone marrow cells (experimental treatment), or control bone marrow cells (control treatment), to recipient bone marrow cells. Six weeks after transplantation, *Pten* was deleted. **b, c**, Two days after pIpC treatment, donor cells accounted for 69 to 71% of Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] HSCs in recipient mice as expected, whether the mice had been transplanted with control (*Pten*^{fl/fl}; *Mx-1-Cre*) donor cells (**b**) or *Pten*^{fl/fl}; *Mx-1-Cre* donor cells (**c**). Values are mean $\pm$ s.d. of three mice per treatment. **d, e**, Subsequently, 5 to 20 weeks after pIpC treatment, control cells still accounted for 64% of bone marrow Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] HSCs (**d**), but *Pten*^{fl/fl}; *Mx-1-Cre* donor cells accounted for only 15% of Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] HSCs (**e**). Values are mean $\pm$ s.d. of ten mice per treatment. *, $P < 0.05$. **f, g**, Control HSCs were stable over time (**f**), whereas *Pten*^{fl/fl}; *Mx-1-Cre* donor HSCs became depleted relative to recipient (wild-type) HSCs in the same mice (**g**). Values are mean $\pm$ s.d.

Consistent with a cell-autonomous requirement for *Pten* in HSCs, the number of *Pten*-deficient HSCs declined over time and the number of wild-type recipient HSCs in the same mice increased. By 5 to 20 weeks after pIpC administration, control *Pten*^{fl/fl}; *Mx-1-Cre* donor cells still accounted for 64% of bone marrow Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells (Fig. 3d), but *Pten*^{fl/fl}; *Mx-1-Cre* donor cells accounted for only 15% of bone marrow Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells (Fig. 3e). The total number of control *Pten*^{fl/fl}; *Mx-1-Cre* donor HSCs was stable over time (Fig. 3f). In contrast, *Pten*^{fl/fl}; *Mx-1-Cre* donor HSCs initially dominated the HSC pool, but by 5 to 20 weeks after pIpC treatment recipient HSCs outnumbered *Pten*-deficient donor HSCs by 6.4-fold (Fig. 3g). To confirm functionally the depletion of donor HSCs, we transplanted 1×10^6 bone marrow cells, or 100 Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells, into irradiated mice. All recipients of bone marrow cells from control mice (from Fig. 3d) showed long-term multilineage reconstitution by donor cells (Supplementary Table 4). However, recipients of *Pten*-deficient donor (from Fig. 3e) bone marrow cells or Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells never achieved multilineage donor cell reconstitution (Supplementary Table 4). As *Pten*-deficient HSCs were depleted and control HSCs were expanded within the same mice, these data confirm that *Pten* is required cell-autonomously for the maintenance of HSCs.

Pten deletion seemed to deplete HSCs by inhibiting self-renewal rather than by promoting cell death. We did not detect any increase in cell death, as assessed by staining for annexin V or activated caspase-3, in either whole bone marrow cells or Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells from *Pten*^{fl/fl}; *Mx-1-Cre* mice four weeks after pIpC treatment (Supplementary Fig. 4). Moreover, approximately 90% of single Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells from either *Pten*-deleted mice or control mice formed colonies in methylcellulose irrespective of whether they were isolated five days or four weeks after pIpC treatment (Supplementary Fig. 5). If HSCs were destined to undergo cell death or rapid terminal differentiation after *Pten* deletion, then Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells from *Pten*-deleted mice should have formed fewer colonies in methylcellulose. Together with the ability of *Pten*-deficient HSCs to efficiently engraft and transiently reconstitute irradiated mice (Fig. 2), these data suggest that HSCs show less self-renewal potential after *Pten* deletion.

Pten-deficient leukaemias are transplantable

Leukaemia-initiating cells are defined by their ability to transfer disease upon transplantation into irradiated mice^{5–7}. These cells are rare among unfractionated leukaemia cells but are highly enriched among cells that express HSC markers^{5–7}. To test whether the neoplasms in *Pten*^{fl/fl}; *Mx-1-Cre* mice were transplantable, we transplanted bone marrow cells, splenocytes, Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells^{25–27}, Mac-1⁺ CD4[−] CD45^{hi} myeloid blasts, Mac-1⁺ B220[−] CD3[−] myeloid cells, CD4⁺ Mac-1[−] CD45^{hi} lymphoid blasts, or CD3⁺ Mac-1[−] B220⁺ Mac-1[−] lymphoid cells from five independent *Pten*^{fl/fl}; *Mx-1-Cre* donors (which were euthanized owing to illness) into irradiated recipients. Virtually every recipient of 5×10^5 to 2×10^6 *Pten*^{fl/fl}; *Mx-1-Cre* donor bone marrow cells (Fig. 4a) or splenocytes (Fig. 4c) died within four weeks of transplantation with ALL and/or AML. Only a minority of the recipients of 3×10^5 *Pten*^{fl/fl}; *Mx-1-Cre* donor bone marrow cells died, with 2 out of 14 developing AML and 2 out of 14 developing ALL (Fig. 4b). By limiting dilution statistics²⁸, this suggests that approximately 1 out of every 600,000 bone marrow cells (0.00017%) used in these experiments were capable of initiating AML or ALL. Recipients of control bone marrow cells ($n = 20$), splenocytes ($n = 7$) or HSCs ($n = 17$) from *Pten*^{fl/fl}; *Mx-1-Cre* mice never developed leukaemia (data not shown).

To test whether leukaemia-initiating cells co-purify with HSCs, we transplanted 10 to 15 Flk-2⁺ Sca-1⁺ Lin[−] c-Kit⁺ CD48[−] cells from *Pten*^{fl/fl}; *Mx-1-Cre* donors into 33 irradiated recipients, five of which died from AML within four weeks (Fig. 4d). This suggests that 1 out

of every 81 cells in this population (1.2%) were capable of initiating AML²⁸—a considerable enrichment compared with whole bone marrow, although the vast majority of cells that co-purified with HSCs did not transfer disease.

Half of the recipients of 15,000 to 25,000 myeloid blasts died from AML within four weeks of transplantation (Fig. 4e), suggesting that 1 out of every 36,000 (0.003%) myeloid blast cells initiated AML. Therefore, AML-initiating cells were enriched among blast cells when compared with whole bone marrow, but not nearly as enriched as among cells that expressed HSC markers. Recipients of bulk myeloid cells, lymphoid blasts and bulk lymphoid cells also developed ALL and/or AML, although leukaemia-initiating cells were not as enriched within these populations (Fig. 4f–h). Thus, a variety of cell populations contained leukaemia-initiating cells.

Rapamycin depletes leukaemia-initiating cells

The observation that *Pten* deletion leads to the depletion of normal HSCs but promotes the generation of leukaemia-initiating cells provided a rare distinction between the mechanisms that regulate the maintenance of normal stem cells compared with leukaemia-initiating cells. The PI(3)K pathway is highly branched, but activates the mammalian target of rapamycin (mTOR) among other downstream effectors^{29,30}. mTOR kinase activity is inhibited by the drug rapamycin^{31,32}, and human AMLs and ALLs have been shown to respond to rapamycin^{33–35}. Therefore, we administered rapamycin to *Pten*^{fl/fl}; *Mx-1-Cre* mice to test whether it depleted leukaemia-initiating cells or rescued normal HSC function.

Pten^{fl/fl}; *Mx-1-Cre* mice became overtly ill after pIpC treatment as they developed leukaemias, exhibiting lethargy, ruffling of fur, and hunched posture (Fig. 5a). All three such mice in this experiment died within 3 to 4 weeks of pIpC treatment from AML and ALL (Fig. 5c). In contrast, three mice that were maintained on daily injections of rapamycin (4 mg per kg of body weight) remained healthy and active four weeks after pIpC treatment (Fig. 5b). These rapamycin-treated mice did not show any histological evidence of neoplasm, as the spleens had normal architecture with only focal areas of erythroid-predominant haematopoiesis (Fig. 5c). Daily injections of rapamycin for seven days after pIpC treatment also prevented the decrease in bone marrow cellularity (Fig. 5d) and the increase in spleen cellularity (Fig. 5e) observed in *Pten*^{fl/fl}; *Mx-1-Cre* mice, without significantly affecting these parameters in control mice. Hence, mice maintained on rapamycin immediately after *Pten* deletion did not develop signs of haematopoietic malignancy.

To determine whether rapamycin eliminated leukaemia-initiating cells, we treated *Pten*-deleted mice with vehicle or rapamycin for six weeks and then transplanted graded doses of whole bone marrow cells into irradiated mice (which no longer received rapamycin). Recipients of bone marrow cells from vehicle-treated mice all died in a dose-dependent manner within 20 to 31 days of transplantation (Fig. 5f). In contrast, recipients of bone marrow cells from rapamycin-treated mice remained healthy and never showed signs of leukaemia, irrespective of the dose of cells transplanted (Fig. 5f). This demonstrates that rapamycin inhibits the generation or maintenance of leukaemia-initiating cells.

To test whether rapamycin was effective against established leukaemias, mice that had been transplanted with *Pten*^{fl/fl}; *Mx-1-Cre* bone marrow cells were treated with daily injections of vehicle or rapamycin, beginning 15 weeks after pIpC administration. Although all three vehicle-treated mice died from ALL and/or AML within five weeks, all three rapamycin-treated mice remained overtly healthy (Supplementary Table 5). Almost all recipients of bone marrow cells from a vehicle-treated mouse died in a dose-dependent manner (Fig. 5g). In contrast, most recipients of bone marrow cells from rapamycin-treated mice survived (Fig. 5g). Thus, rapamycin reduced the frequency of leukaemia-initiating cells even when treatment was initiated after the onset of frank leukaemia. The two rapamycin-treated mice that were not sacrificed to provide a source of cells for

transplantation were treated with daily injections of rapamycin for 15 weeks. Although these mice seemed overtly healthy, with normalized spleens and thymuses, one showed histological evidence of myeloproliferative disease and the other showed signs of AML and ALL.

Rapamycin treatment was also initiated after the transplantation of 2×10^6 bone marrow cells from a *Pten*-deficient mouse with AML and ALL into irradiated recipients. Vehicle-treated recipients all died

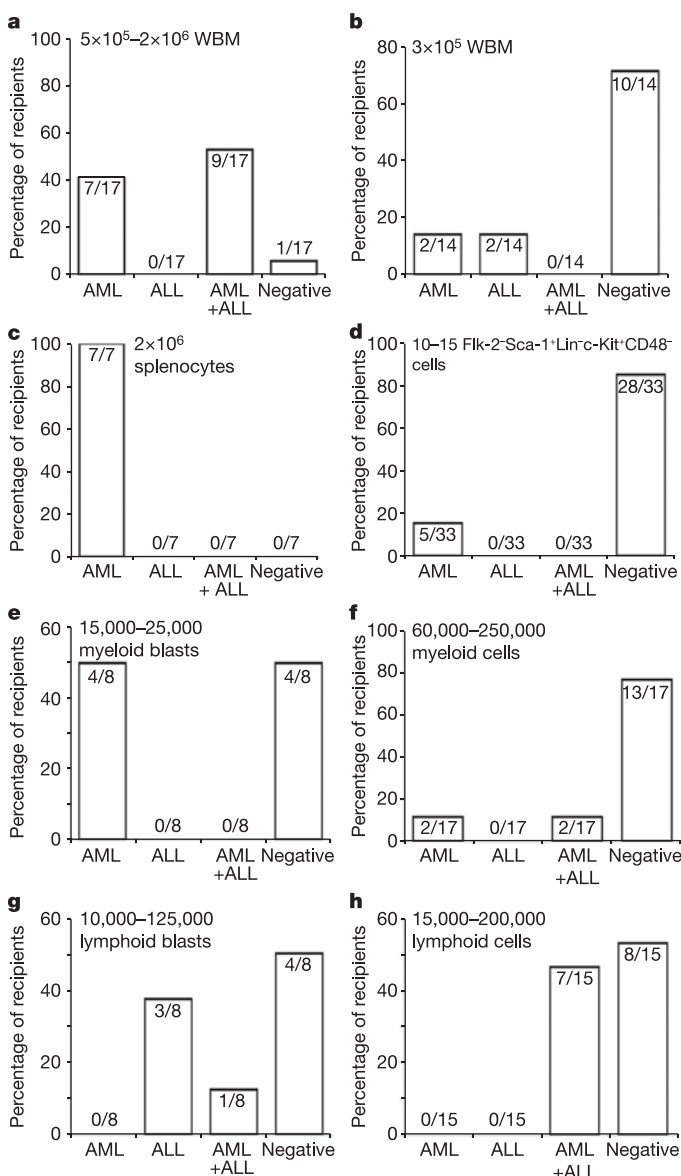


Figure 4 | AML- and ALL-initiating cells are rare in *Pten*-deleted mice, but are transplantable, are contained within multiple distinct populations, and are highly enriched among cells that express HSC markers.

a–h, Whole bone marrow (WBM) cells (**a**, **b**), whole splenocytes (**c**), Flk-2⁺Sca-1⁺Lin[−]c-Kit⁺CD48⁺ HSCs^{25,26} (**d**), CD45^{hi}Mac-1⁺CD4[−] myeloid blasts (**e**), Mac-1⁺B220[−]CD3[−] myeloid cells (**f**), CD45^{hi}CD4⁺Mac-1[−] lymphoid blasts (**g**), or CD3⁺Mac-1[−]/B220⁺Mac-1[−] lymphoid cells (**h**) were transplanted into irradiated recipients from 3 to 5 independent *Pten*^{fl/fl}; *Mx-1-Cre* donor mice with leukaemia. Donor (CD45.2⁺) cells were competed against 200,000 recipient (CD45.1⁺) bone marrow cells for radioprotection. Bars indicate the proportion of recipients that died with AML, ALL, or AML and ALL, or which survived with no signs of neoplasia ('Negative'). Recipients of *Pten*^{fl/+}; *Mx-1-Cre* control bone marrow cells ($n = 20$), splenocytes ($n = 7$) or HSCs ($n = 17$) never developed leukaemia (data not shown).

within 25 days of transplantation (Fig. 5h). In contrast, rapamycin-treated recipients died 40 to 60 days after transplantation (Fig. 5h). When initiated after the onset of leukaemia, rapamycin was effective in prolonging the life of mice, but not in curing the leukaemias.

Rapamycin also inhibited the survival and proliferation of clonogenic leukaemia cells in culture. Freshly isolated or cultured myeloid blast cells from *Pten^{fl/fl}; Mx-1-Cre* mice with AML were plated into methylcellulose. Rapamycin significantly reduced the percentage of blast cells that formed colonies, as well as colony size, in a dose-dependent manner (Supplementary Fig. 6a–e). Rapamycin also significantly reduced the percentage of myeloid blasts in S phase of the cell cycle, and increased the percentage of cells expressing activated caspase-3 (Supplementary Fig. 6f, g).

Rapamycin rescues *Pten*-deficient HSCs

Rapamycin also restored the capacity of *Pten*-deficient HSCs to provide long-term multilineage reconstitution to irradiated mice. Daily injections of rapamycin for seven days after pIpC administration did not affect the overall rate of proliferation in bone marrow, but did normalize the cell cycle distribution of Flk-2⁺Sca-1⁺Lin[−]c-Kit⁺CD48[−] cells in *Pten^{fl/fl}; Mx-1-Cre* mice without affecting the proliferation of HSCs from control littermates (Fig. 6a). Rapamycin also eliminated the HSC expansion observed seven days after pIpC treatment (Fig. 6b), and the HSC depletion observed after four weeks in *Pten^{fl/fl}; Mx-1-Cre* mice (Fig. 6c), without affecting HSC numbers in control mice. Most notably, rapamycin restored the potential of Flk-2⁺Sca-1⁺Lin[−]c-Kit⁺CD48[−] cells isolated from

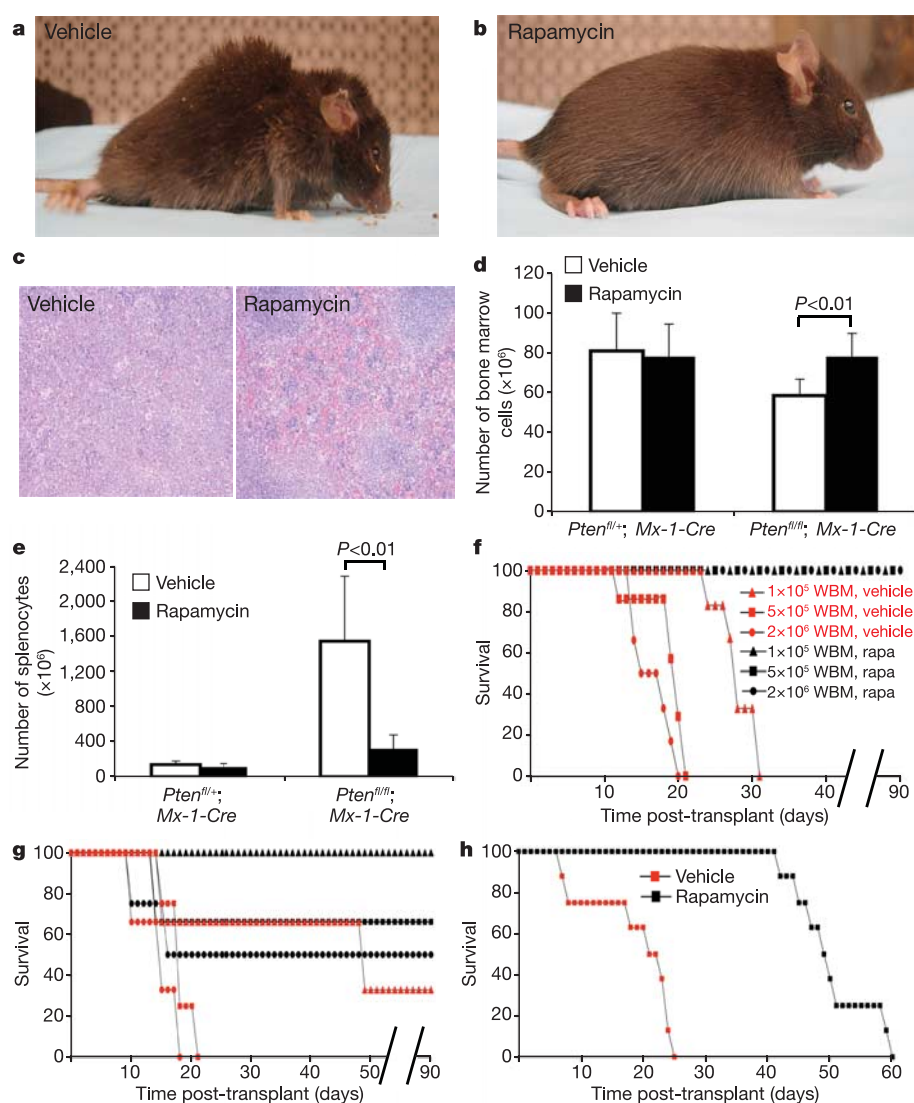


Figure 5 | Rapamycin depletes leukaemia-initiating cells. **a, b**, Four weeks after pIpC treatment, *Pten^{fl/fl}; Mx-1-Cre* mice exhibited lethargy, ruffled fur, hunched posture, and required euthanasia (**a**; $n = 3$), but if treated daily with 4 mg per kg rapamycin remained healthy and active (**b**; $n = 3$). **c**, Rapamycin-treated mice had normal spleen architecture (right panel), whereas vehicle-treated littermates developed AML and ALL (left panel). **d, e**, Rapamycin prevented the decrease in bone marrow cellularity (**d**) and the increase in spleen cellularity (**e**) in *Pten^{fl/fl}; Mx-1-Cre* mice, but did not affect these parameters in controls. Values are mean $\pm$ s.d. of 6 to 7 mice per treatment. **f**, Mice transplanted with 1×10^6 *Pten^{fl/fl}; Mx-1-Cre* donor (CD45.2) bone marrow cells along with 0.5×10^6 control (CD45.1) bone marrow cells were treated with vehicle or rapamycin (4 mg per kg per day) for six weeks, starting immediately after pIpC treatment. Graded doses of

whole bone marrow (WBM) cells from these mice were transplanted into lethally irradiated recipients along with 2×10^5 recipient (CD45.1) bone marrow cells. Recipients of bone marrow from vehicle-treated mice died with leukaemia, whereas recipients of marrow from rapamycin-treated mice remained healthy. **g**, Similar donor mice were treated for five weeks with vehicle or rapamycin (4 mg per kg per day) beginning 15 weeks after pIpC treatment. Nearly all recipients of bone marrow from vehicle-treated mice died of leukaemia in a dose-dependent manner, whereas recipients of cells from rapamycin-treated mice usually survived. **h**, Finally, 2×10^6 donor (CD45.2) bone marrow cells from a *Pten^{fl/fl}; Mx-1-Cre* mouse with AML and ALL were transplanted into sublethally irradiated recipients. The recipients were treated with vehicle or rapamycin (0.4 mg per kg per day). Panels **f** and **h** show one representative experiment out of three.

Pten^{fl/fl}; *Mx-1-Cre* mice to provide long-term multilineage reconstitution seven days after pIpC treatment (Fig. 6d, e). This confirms that Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ cells from *Pten*^{fl/fl}; *Mx-1-Cre* mice are HSCs, and that rapamycin restores normal function to these cells.

The mechanism responsible for the depletion of *Pten*-deficient HSCs remains to be elucidated. One possibility is that persistent activation of the PI(3)K pathway following the loss of *Pten* leads to reduced HSC self-renewal via a gradual increase in the rate at which HSCs exit the stem cell pool. Another possibility is that *Pten* deficiency induces the gradual senescence of HSCs. Conditional deletion of *Pten* leads to a p53-dependent senescence of prostate cells³⁶. Leukaemias might acquire secondary mutations that inactivate the senescence response.

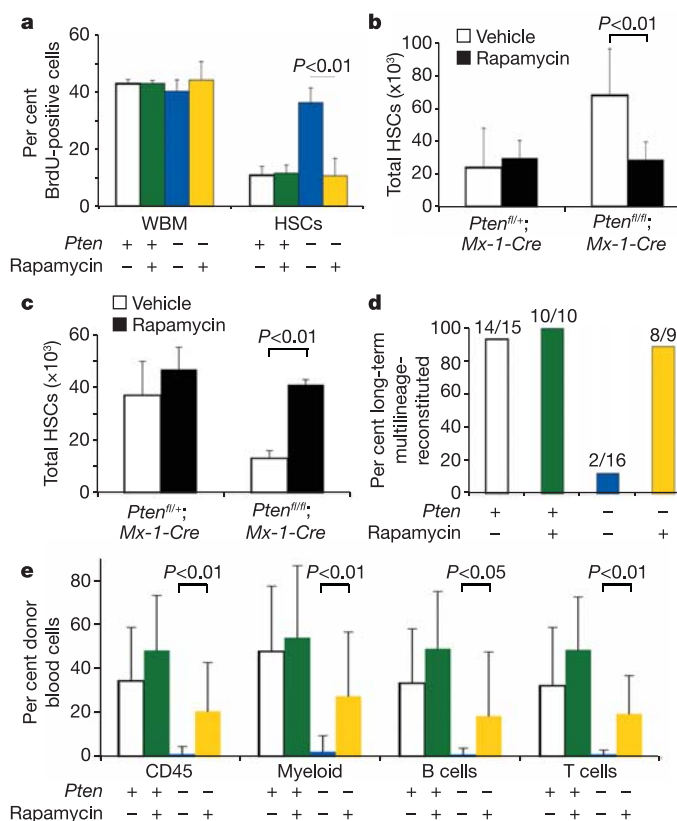


Figure 6 | Rapamycin rescues normal HSC function after *Pten* deletion.

a, Seven days of rapamycin (4 mg per kg per day) did not affect proliferation in whole bone marrow (WBM; 19-h pulse with BrdU), but did eliminate the increase in Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ HSC proliferation in *Pten*^{fl/fl}; *Mx-1-Cre* mice without affecting proliferation in controls. Values are mean $\pm$ s.d. of three experiments. **b**, Rapamycin eliminated the HSC expansion observed seven days after pIpC treatment in *Pten*^{fl/fl}; *Mx-1-Cre* mice without affecting HSCs in *Pten*^{fl/fl}; *Mx-1-Cre* controls. Values are mean $\pm$ s.d. of 6 to 7 mice per treatment. **c**, Rapamycin eliminated the depletion of HSCs observed 3 to 4 weeks after pIpC treatment in *Pten*^{fl/fl}; *Mx-1-Cre* mice without affecting controls. Values are mean $\pm$ s.d. of three mice per treatment. **d**, **e**, Rapamycin restored the long-term (>16 weeks after transplantation) multilineage reconstitution potential of Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ cells from *Pten*^{fl/fl}; *Mx-1-Cre* mice (**d**). Fifteen Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ donor HSCs from *Pten*^{fl/fl}; *Mx-1-Cre* or control mice that had been treated with rapamycin (4 mg per kg per day) or vehicle for seven days after pIpC treatment were transplanted along with 200,000 recipient bone marrow cells into irradiated recipients. Recipients received vehicle or rapamycin (0.4 mg per kg per day) starting within two weeks of transplantation. The frequency of donor white blood (CD45.2⁺) cells, myeloid (Mac-1⁺) cells, B (B220⁺) cells and T (CD3⁺) cells was determined 16 to 18 weeks after transplantation (**e**). Values are mean $\pm$ s.d. of three experiments.

Deletion of the cyclin-dependent kinase inhibitor *p21*^{Cip1} also leads to HSC proliferation followed by a slow depletion³⁷. Like PTEN, *p21*^{Cip1} also regulates HSC quiescence. This raises the possibility that compounds that promote stem cell quiescence might consistently have different effects on normal stem cells and cancer stem cells.

Conditional *Pten* deletion in the fetal central nervous system increases the self-renewal and frequency of neural stem cells³⁸. This is the opposite of what we observed, and may represent a general difference between fetal and adult stem cells rather than a difference between tissues. The balance of proto-oncogenes and tumour suppressor genes that regulate stem cell self-renewal changes between embryonic, fetal and adult life as the organogenic demand decreases and the risk of cancer increases^{11,13,39}.

These data demonstrate that it is possible to identify—and to target therapeutically—pathways that have distinct effects on normal stem cells and cancer stem cells within the same tissue. This is an important finding, because it has been proposed that oncogenic mutations confer self-renewal potential by activating pathways used by normal stem cells, irrespective of whether the mutations occur in stem cells or other cells^{1,2,12}. By comparing the mechanisms that regulate the maintenance of normal stem cells versus cancer stem cells, it should be possible to design new therapies and to improve existing therapies.

METHODS

Mice. Mice were housed in the Unit for Laboratory Animal Medicine at the University of Michigan. *Pten*^{fl/fl} and *Mx-1-Cre* mice were backcrossed for eight and six generations, respectively, onto a C57BL/Ka-CD45.2:Thy-1.1 background. Recipients in reconstitution assays were adult C57BL/Ka-CD45.1:Thy-1.2 mice.

Flow cytometry and isolation of HSCs. Bone marrow cells were flushed from the long bones (tibias and femurs) with Hank's buffered salt solution without calcium or magnesium, supplemented with 2% heat-inactivated calf serum (HBSS⁺; GIBCO). Cells were triturated and filtered through nylon screen (45 μ m; Sefar America) to obtain a single-cell suspension. Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ HSCs and Thy-1.1^{low}Sca-1⁺Mac-1^{low}CD4^{low}B220⁻ multipotent progenitors (MPPs) were isolated as previously described^{25,26,40}. For isolation of Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ HSCs, whole bone marrow cells were incubated with unconjugated monoclonal antibodies to lineage (Lin) markers including B220 (6B2), CD3 (KT31.1), CD4 (GK1.5), CD5 (53-7.3), CD8 (53-6.7), Gr-1 (8C5), Mac-1 (M1/70) and Ter119. After washing, cells were resuspended in anti-rat IgG conjugated to phycoerythrin (PE; Jackson ImmunoResearch). Cells were then stained with directly conjugated antibodies to Sca-1 (Ly6A/E-APC), c-Kit (2B8-biotin), Flk-2 (A2F10-PE; eBioscience) and CD48 (HM48-1-FITC; BD Pharmingen). To identify CD45.2⁺ HSCs, antibodies to CD45.2 (104-FITC; BD Pharmingen) and CD48 (HM48-1-PE; eBioscience) were used. HSCs were often pre-enriched by selecting c-Kit⁺ cells using paramagnetic microbeads and autoMACS (Miltenyi Biotec). To identify leukaemic blast cells, anti-CD45 (30-F11-APC; eBioscience) was used.

The total number of Flk-2⁺Sca-1⁺Lin⁻c-Kit⁺CD48⁻ cells per mouse was calculated based on the frequency of this population in the bone marrow and spleen, the cellularity of the spleen and long bones, and by assuming that 15% of all bone marrow is within the long bones⁴¹. The blood and other tissues do not contribute significantly to the overall size of the HSC pool.

Long-term competitive reconstitution assays. Adult recipient mice were irradiated with an Orthovoltage X-ray source delivering 300 rad min⁻¹. Recipient mice received two doses of 540 rads each, delivered 3 h apart. For sublethal irradiation, mice were administered one dose of 800 rads. Donor (CD45.2⁺) HSCs were sorted and then re-sorted (for purity) into individual wells of a 96-well plate containing 200,000 CD45.1⁺ whole bone marrow cells in HBSS⁺. The contents of individual wells were injected into the retro-orbital venous sinus of irradiated CD45.1⁺ recipients. For at least 16 weeks after transplantation, blood was obtained from the tail veins of recipient mice, subjected to ammonium chloride/potassium bicarbonate red-cell lysis, and stained with directly conjugated antibodies to CD45.2 (104-FITC), B220 (6B2), Mac-1 (M1/70), CD3 (KT31.1) and Gr-1 (8C5) to assess donor cell engraftment. Cell cycle analysis was conducted as described in Supplementary Methods.

Administration of pIpC and rapamycin. As described previously⁴², polyinosine-polycytidine (pIpC; Sigma) was resuspended in Dulbecco's-PBS at 2 mg ml⁻¹ and passed through a 0.22- μ m filter. Mice received 25 μ g of pIpC per gram of body mass every other day for two weeks. Rapamycin (Calbiochem and LC Laboratories) was administered by intraperitoneal injection at the indicated

doses. It was reconstituted in absolute ethanol at 10 mg ml^{-1} or 1 mg ml^{-1} and diluted in 5% Tween-80 (Sigma) and 5% PEG-400 (Hampton Research) before injection. The final volume of all injections was $200 \mu\text{l}$.

Received 1 October 2005; accepted 1 March 2006.

Published online 5 April 2006.

- Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
- Pardal, R., Clarke, M. F. & Morrison, S. J. Applying the principles of stem-cell biology to cancer. *Nature Rev. Cancer* **3**, 895–902 (2003).
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100**, 3983–3988 (2003).
- Singh, S. K. *et al.* Identification of a cancer stem cell in human brain tumors. *Cancer Res.* **63**, 5821–5828 (2003).
- Lapidot, T. *et al.* A cell initiating human acute myeloid leukemia after transplantation into SCID mice. *Nature* **17**, 645–648 (1994).
- Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Med.* **3**, 730–737 (1997).
- Warner, J. K., Wang, J. C., Hope, K. J., Jin, L. & Dick, J. E. Concepts of human leukemic development. *Oncogene* **23**, 7164–7177 (2004).
- Park, I.-K. *et al.* Bmi-1 is required for the maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302–305 (2003).
- Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255–260 (2003).
- Hemmati, H. D. *et al.* Cancerous stem cells can arise from pediatric brain tumors. *Proc. Natl Acad. Sci. USA* **100**, 15178–15183 (2003).
- Molofsky, A. V., Pardal, R. & Morrison, S. J. Diverse mechanisms regulate stem cell self-renewal. *Curr. Opin. Cell Biol.* **16**, 700–707 (2004).
- Taipale, J. & Beachy, P. A. The Hedgehog and Wnt signalling pathways in cancer. *Nature* **411**, 349–354 (2001).
- Molofsky, A. V. *et al.* Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the p16^{Ink4a} and p19^{Arf} senescence pathways. *Genes Dev.* **19**, 1432–1437 (2005).
- Lowe, S. W. & Sherr, C. J. Tumor suppression by *Ink4a-Arf*: progress and puzzles. *Curr. Opin. Genet. Dev.* **13**, 77–83 (2003).
- Maehama, T. & Dixon, J. E. The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate. *J. Biol. Chem.* **273**, 13375–13378 (1998).
- Stiles, B., Groszer, M., Wang, S., Jiao, J. & Wu, H. PTENless means more. *Dev. Biol.* **273**, 175–184 (2004).
- Di Cristofano, A. & Pandolfi, P. P. The multiple roles of PTEN in tumor suppression. *Cell* **100**, 387–390 (2000).
- Aggerholm, A., Gronbaek, K., Guldberg, P. & Hokland, P. Mutational analysis of the tumour suppressor gene MMAC1/PTEN in malignant myeloid disorders. *Eur. J. Haematol.* **65**, 109–113 (2000).
- Roman-Gomez, J. *et al.* Promoter hypermethylation of cancer-related genes: a strong independent prognostic factor in acute lymphoblastic leukemia. *Blood* **104**, 2492–2498 (2004).
- Dahia, P. L. *et al.* PTEN is inversely correlated with the cell survival factor Akt/PKB and is inactivated via multiple mechanisms in haematological malignancies. *Hum. Mol. Genet.* **8**, 185–193 (1999).
- Cheong, J. W. *et al.* Phosphatase and tensin homologue phosphorylation in the C-terminal regulatory domain is frequently observed in acute myeloid leukaemia and associated with poor clinical outcome. *Br. J. Haematol.* **122**, 454–456 (2003).
- Hock, H. *et al.* Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival. *Genes Dev.* **18**, 2336–2341 (2004).
- Kühn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
- Kogan, S. C. *et al.* Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice. *Blood* **100**, 238–245 (2002).
- Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl Acad. Sci. USA* **98**, 14541–14546 (2001).
- Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Yilmaz, O. H., Kiel, M. J. & Morrison, S. J. SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity. *Blood* **107**, 924–930 (2005).
- Smith, L. G., Weissman, I. L. & Heimfeld, S. Clonal analysis of hematopoietic stem-cell differentiation *in vivo*. *Proc. Natl Acad. Sci. USA* **88**, 2788–2792 (1991).
- Majumder, P. K. *et al.* mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. *Nature Med.* **10**, 594–601 (2004).
- Inoki, K., Corradetti, M. N. & Guan, K. L. Dysregulation of the TSC–mTOR pathway in human disease. *Nature Genet.* **37**, 19–24 (2005).
- Podsypanina, K. *et al.* An inhibitor of mTOR reduces neoplasia and normalizes p70/S6 kinase activity in *Pten*^{+/-} mice. *Proc. Natl Acad. Sci. USA* **98**, 10320–10325 (2001).
- Neshat, M. S. *et al.* Enhanced sensitivity of PTEN-deficient tumors to inhibition of FRAP/mTOR. *Proc. Natl Acad. Sci. USA* **98**, 10314–10319 (2001).
- Recher, C. *et al.* Antileukemic activity of rapamycin in acute myeloid leukemia. *Blood* **105**, 2527–2534 (2005).
- Avellino, R. *et al.* Rapamycin stimulates apoptosis of childhood acute lymphoblastic leukemia cells. *Blood* **106**, 1400–1406 (2005).
- Teachey, D. T. *et al.* The mTOR inhibitor CCI-779 induces apoptosis and inhibits growth in preclinical models of primary adult human ALL. *Blood* **107**, 1149–1155 (2006).
- Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of Pten-deficient tumorigenesis. *Nature* **436**, 725–730 (2005).
- Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21cip1/waf1. *Science* **287**, 1804–1808 (2000).
- Groszer, M. *et al.* Negative regulation of neural stem/progenitor cell proliferation by the *Pten* tumor suppressor gene *in vivo*. *Science* **294**, 2186–2189 (2001).
- Molofsky, A. V. *et al.* Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962–967 (2003).
- Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673 (1994).
- Smith, L. H. & Clayton, M. L. Distribution of injected 59Fe in mice. *Exp. Hematol.* **20**, 82–86 (1970).
- Mikkola, H. K. *et al.* Haematopoietic stem cells retain long-term repopulating activity and multipotency in the absence of stem-cell leukaemia *SCL/tal-1* gene. *Nature* **421**, 547–551 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Howard Hughes Medical Institute. O.H.Y. was supported by a predoctoral fellowship from the University of Michigan (UM) Institute of Gerontology. We thank the UM Flow-cytometry Core Facility, which was supported by the UM-Comprehensive Cancer Center. We also thank E. Smith in the Hybridoma Core Facility for antibody production, supported in part through the Rheumatic Disease Core Center; A. Burgess and N. McAnsh of the UM Comprehensive Cancer Center Tissue Core; and C. Mountford for excellent mouse colony management.

Author Contributions O.H.Y. performed all experiments and participated in the design and interpretation of experiments. R.V. performed all pathology on the mice with help from O.H.Y. B.K.T. and D.O.F. performed spectral karyotype analysis with help from O.H.Y. W.G. and H.W. provided the *Pten*^{fl/fl} mice and discussed pre-publication results. S.J.M. participated in the design and interpretation of experiments, and wrote the paper with O.H.Y.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.J.M. (seanjm@umich.edu).

Chondrule formation in particle-rich nebular regions at least hundreds of kilometres across

Jeffrey N. Cuzzi¹ & Conel M. O'D. Alexander²

Chondrules are millimetre-sized spherules (mostly silicate) that dominate the texture of primitive meteorites¹. Their formation mechanism is debated, but their sheer abundance suggests that the mechanism was both energetic and ubiquitous in the early inner Solar System². The processes suggested—such as shock waves, solar flares or nebula lightning^{3–7}—operate on different length scales that have been hard to relate directly to chondrule properties. Chondrules are depleted in volatile elements, but surprisingly they show little evidence for the associated loss of lighter isotopes one would expect⁸. Here we report a model in which molten chondrules come to equilibrium with the gas that was evaporated from other chondrules, and which explains the observations in a natural way. The regions within which the chondrules formed must have been larger than 150–6,000 km in radius, and must have had a precursor number density of at least 10 m^{-3} . These constraints probably exclude nebula lightning, and also make formation far from the nebula midplane problematic. The wide range of chondrule compositions may be the result of different combinations of the local concentrations of precursors and the local abundance of water ice or vapour.

The ideal situation of evaporation into vacuum was first described by Rayleigh⁹, and has since been validated by numerous experiments¹⁰. Under these conditions, lighter isotopes in the evaporating (volatile) species have higher thermal velocities, so the residual material becomes preferentially depleted in the lighter isotopes, or isotopically fractionated. Although chondrules are depleted in volatile elements like K, Fe, Si and Mg (refs 8 and 11), the systematic isotopic fractionations associated with evaporation are not seen. There are several hypotheses for why chondrules might show no isotopic fractionations: (1) chondrules came to equilibrium with the gas evaporated from other chondrules, (2) the ambient gas pressure was extremely high, restricting the diffusion of evaporated vapour away from the chondrule, or (3) chondrules were heated and cooled so rapidly that little evaporation occurred. Experiments that simulate chondrule textures tend to exclude case (3) (ref. 12). Here we show that case (2) is implausible, and present a simple physical model for case (1) that specifies the concentrations and associated size scales for chondrule formation regions needed to reproduce the observations.

The net evaporation rate for a single chondrule is given by:

$$\dot{m}_{\text{evap}} = \pi r_c^2 \gamma (\rho_{\text{sat}} - \rho_x) v_{\text{th}} \quad (1)$$

where r_c is the chondrule radius, γ is an evaporation coefficient, ρ_{sat} is the saturation vapour mass density, ρ_x is the local vapour density, and v_{th} is the thermal velocity of evaporated atoms or molecules (see Supplementary Information). If chondrules can reach equilibrium with the evaporated gas, they can avoid preserving the initial isotopic fractionation which accompanies this evaporation^{13,14}. At equilibrium, the ambient vapour density ρ_x must equal the saturation vapour density ρ_{sat} .

We discuss two limiting cases leading to this situation¹⁵. Case (1) allows expansion of evaporated vapour into overlapping clouds much larger than the mean chondrule spacing, so that chondrules are bathed in the evaporated material from many other chondrules. Expansion of the evaporated vapour cloud is limited by diffusion in the ambient hydrogen gas, reaching some characteristic size $R(t)$ of the order of $(Dt)^{1/2}$ in time t , where $D = 1.4 \times 10^{-4} T^{3/2} / P$ is the pressure-dependent diffusion coefficient¹⁰, and P is pressure in bar.

The local density $\rho_x(t)$ due to overlapping contributions from all evaporating chondrules, at all prior times, can be obtained by summing over all diffusion profiles of all surrounding chondrules as (see Supplementary Information):

$$\rho_x(t) = \int_0^{R_1} 4\pi R^2 n_c dR \int_0^t \frac{\dot{m}_{\text{evap}}(t')}{(4\pi Dt')^{3/2}} e^{-R^2/4Dt'} dt' \quad (2)$$

where R_1 is the radius of the region of chondrule precursors about the local point, having constant volume density n_c , and we use $\dot{m}_{\text{evap}}$ from equation (1). If $R_1 > (3-4)\sqrt{Dt}$ (as seen from the point x) the radial integral is known and the equation simplifies to:

$$\rho_x(t) = n_c \pi r_c^2 \gamma v_{\text{th}} \int_0^t (\rho_{\text{sat}} - \rho_x(t')) dt' \quad (3)$$

which, for heating time $t = t_h$, has the solution $\rho_x(t) = \rho_{\text{sat}}(1 - e^{n_c \pi r_c^2 \gamma v_{\text{th}} t_h}) = \rho_{\text{sat}}(1 - e^{-a})$, where

$$a = n_c \pi r_c^2 \gamma v_{\text{th}} t_h = (A\Phi)(3/4\rho_s r_c) \rho_g \gamma v_{\text{th}} t_h \quad (4)$$

The second equality relates a to the local chondrule-precursor enrichment factor $\Phi = 4\pi\rho_s r_c^3 n_c / 3A\rho_g = 4ar_c\rho_s / 3A\rho_g \gamma v_{\text{th}} t_h$ relative to hydrogen by mass, given the silicate cosmic abundance $A \approx 5 \times 10^{-3}$, the chondrule material density ρ_s , and the local gas density ρ_g . How closely ρ_x approaches ρ_{sat} (the value of a) depends not only on the observed isotopic fractionations in the chondrules, but also on the fraction of material in the vapour compared to that in the chondrules¹⁶; values of $a = 3-7$ lead to $\rho_x = (0.95-0.999)\rho_{\text{sat}}$.

In case (2), evaporating material is constrained by very slow diffusion in a high local H_2 (total) gas pressure to a sufficiently small region of radius R_2 , so that its vapour density reaches ρ_{sat} in the heating time t_h ; in this case R_2 can be much smaller than the mean chondrule spacing, and the chondrule merely equilibrates with its own vapour cloud^{17,18}. Pressures required to produce such a situation are of the order of 10 bar (see Supplementary Information). However, canonical nebula pressures in the asteroid belt region are in the range of 10^{-6} to 10^{-3} bar (depending on model assumptions), which is far too small to produce this effect. We thus consider case (1) to be the most likely option.

The second equality of equation (4) may be used to scale our model to the results of chemical kinetic models¹¹. In Fig. 1, we show that the simple model explains and unifies numerical results over a wide variety of conditions. The first equality of equation (4) can then

¹Space Science Division, Ames Research Center, Moffett Field, California 94035, USA. ²Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA.

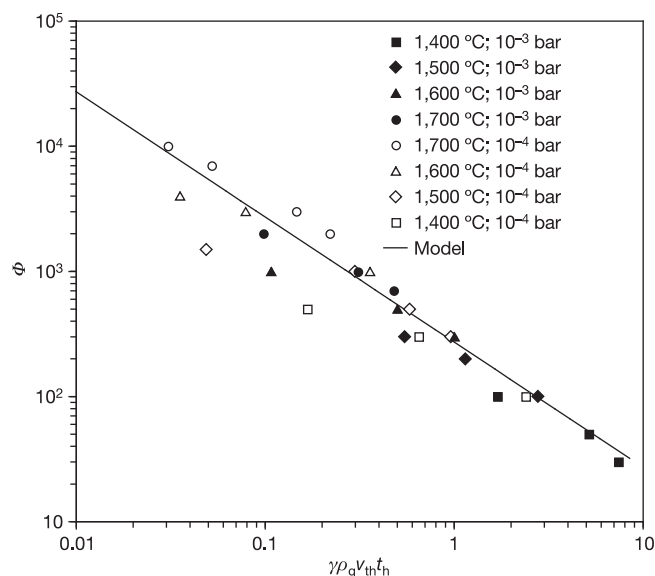


Figure 1 | The mass density in clumps where chondrules form is captured in our model by a simple combination of local properties. The mass enrichment factor Φ over cosmic abundance (equation (4)) required to forestall Mg isotopic fractionation ($\leq 0.25\%$ per atomic mass unit), is plotted as a function of the combined variable $\rho_g \gamma v_{th} t_h$. The line is the second equality of equation (4), and the points are results from numerical models of chemical kinetics¹¹. Our simple model captures the form of the dependence of Φ on the combined local quantity $\rho_g \gamma v_{th} t_h$ very well, with numerical results for pressures of both 10^{-3} and 10^{-4} bar collapsing into the same trend; we require $a = 6 \pm 1$ for a quantitative match. Refinements beyond the scope of this paper greatly improve the fit; see the Supplementary Information. An alternative, but less often used, approach to calculating chondrule evaporation in non-equilibrium situations leads to an equally good fit, but the values of γ are roughly three times as large for this model, which would decrease n_c accordingly. Symbols show results of ref. 11 for different combinations of nebula temperature and pressure.

be solved for n_c for any combination of parameters. Additionally assuming ρ_g , we can solve for Φ . For instance, $a = 6$, $\gamma \approx 0.05$, $r_c = 0.03\text{--}0.05$ cm, and chondrule heating conditions¹⁹ characterized by $T = 1,770$ K for $t_h = 2 \times 10^4$ s lead to $n_c = 18\text{--}7$ m⁻³ respectively. Comparable, or slightly lower, chondrule densities have been derived from other chondrule properties^{3,4}, as discussed in the Supplementary Information. If the clump size is smaller than $(3\text{--}4)\sqrt{D t_h}$, the values of n_c must be larger.

These conditions characterize the chondrule formation regions. For example, in the context of a shock-wave chondrule-heating model, both gas and evaporating chondrules have been compressed by about an order of magnitude in the post-shock region^{3,20}. If the pre-shock gas density is a typical minimum mass nebula value $\rho_g \approx 10^{-10}$ g cm⁻³, then $\Phi \approx 1,400\text{--}2,300$ for $r_c = 0.03\text{--}0.05$ cm, $a = 6$, and $\rho_s = 3.5$ g cm⁻³. Higher pre-shock gas densities^{3,4} of $\rho_g \approx 10^{-9}$ g cm⁻³ are found in disk models with mass accretion rates typical of classical T Tauri stars, and lead to smaller values of $\Phi \approx 140\text{--}230$. Simple ‘settling to the midplane’ is unlikely to satisfy all the relevant criteria, but concentration of the chondrule precursors in nebula turbulence might do so²¹, as discussed further in the Supplementary Information.

The above results imply that at least $10^9\text{--}10^{12}$ chondrules form simultaneously in a volume of radius $R_1 \geq (3\text{--}4)\sqrt{D t_h} \approx 0.5\text{--}20$ km (assuming $T = 1,770$ K, $t_h = 2 \times 10^4$ s, and local gas pressure of $10^{-3}\text{--}10^{-5}$ bar). Chondrules in this co-formed batch equilibrate chemically with each other within t_h , through rapid exchange of their evaporated atoms. If more than 99% of all chondrules in a region of radius R are to escape significant isotopic fractionation, then $(1 - R_1/R)^3 > 0.99$; thus $R > 300R_1 \approx 150\text{--}6,000$ km (or larger values of n_c are needed).

This length scale distinguishes between competing models of chondrule formation. Nebula lightning, for instance, affects a region no more than a few kilometres across⁷. Bow shocks from planetesimals are comparable in size to the planetesimal radius, and thus would also be much smaller than our inferred length scales, except in the unlikely event that most planetesimal growth had already been accomplished^{4,22}. If chondrule formation regions at least $150\text{--}6,000$ km in scale are needed to satisfy the isotopic fractionation constraints, these smaller-scale mechanisms can be excluded independently of the usual arguments based on heating and cooling timescales, which are subject to different uncertainties. Nebula shocks due to gravitational instabilities or jovian density waves are on far larger scales, comparable to the nebula thickness^{3,20}; here the effective spatial scale of the chondrule-forming event is provided by the spatial scale of the dense cloud of precursors, which is amenable to modelling²¹. However, fast shocks from solar flares⁶ act primarily at low ambient nebula densities of $\sim 5 \times 10^{-12}$ to 5×10^{-11} g cm⁻³, requiring chondrule precursor concentration factors to be 20–200 times larger, and occur over volumes $>10\text{--}30$ times larger in radius (because of the dependence of D on density) than for $\rho_g \approx 10^{-9}$ g cm⁻³.

Finally, we note that these results have broader implications. A long-standing concern with nebula melting scenarios has been that chondrule melts are not stable given cosmic abundances; however, the conditions we derive here to preclude isotopic fractionation also stabilize silicate melts^{11,14,16}. Moreover, the compositions of these melts will reflect local conditions. For instance, it is well known that most chondrules in ordinary and carbonaceous chondrites are too oxidized to have formed in a canonical solar mixture¹⁶, requiring local oxygen abundances that are orders of magnitude larger than nominal. Such oxygen enrichments would occur if $\Phi \gg 1$ and the chondrule precursor materials were FeO-rich^{11,16}. Addition of water can also affect oxidation state; this could occur either if chondrule precursors contained water ice, or if chondrules formed sunward of water-ice evaporation fronts^{23,24} where water was enriched in the vapour. Water is likely to carry large oxygen-isotopic anomalies produced by ultraviolet photodissociation in the outer nebula or protostellar cloud^{25–27}. Consequently, local variations in the chondrule precursor concentration factor would produce a range of possibly correlated elemental, isotopic and redox properties even amongst chondrule batches formed in the same event. More extensive observations could constrain the small fraction of chondrules which do show isotopic fractionations. The wide range of chondrule compositions, an important piece of the chondrule puzzle, may be caused at least partially by different combinations of the local concentration of chondrule precursors and global effects such as evaporation fronts. Finally, although this is beyond the scope of this paper, the model outlined above could also be used to constrain the formation conditions of the enigmatic calcium-aluminium-rich inclusions (CAIs) that do preserve variable degrees of isotopic fractionation^{8,10}.

Received 5 January; accepted 17 April 2006.

1. Jones, R. H., Grossman, J. N. & Rubin, A. E. in *Chondrules and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) Vol. 341, 251–316 (Astron. Soc. Pacif. Conf. Ser., San Francisco, 2005).
2. Ciesla, F. in *Chondrules and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) Vol. 341, 811–820 (Astron. Soc. Pacif. Conf. Ser., San Francisco, 2005).
3. Desch, S. & Connolly, H. C. Jr. A model of the thermal processing of particles in solar nebula shocks; application to the cooling rates of chondrules. *Meteor. Planet. Sci.* **37**, 183–207 (2002).
4. Ciesla, F. J., Hood, L. L. & Weidenschilling, S. Evaluating planetesimal bow shocks as sites for chondrule formation. *Meteor. Planet. Sci.* **39**, 1809–1821 (2004).
5. Iida, A., Nakamoto, T., Susa, H. & Nakagawa, Y. A shock heating model for chondrule formation in a protoplanetary disk. *Icarus* **153**, 430–450 (2001).
6. Nakamoto, T., Hayashi, M. R., Kida, N. T. & Tachibana, S. in *Chondrules and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) Vol. 341, 883–892 (Astron. Soc. Pacif. Conf. Ser., San Francisco, 2005).
7. Desch, S. & Cuzzi, J. N. Formation of chondrules by lightning in the protoplanetary nebula. *Icarus* (PPIV special issue) **143**, 87–105 (2000).

8. Davis, A. M., Alexander, C. M. O'D., Nagahara, H. & Richter, F. M. in *Chondrules and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) Vol. 341, 432–455 (Astron. Soc. Pacific Conf. Ser., San Francisco, 2005).
9. Rayleigh, J. W. S. Theoretical considerations respecting the separation of gases by diffusion and other processes. *Phil. Mag.* **42**, 493–498 (1896).
10. Richter, F. M., Davis, A. M., Ebel, D. S. & Hashimoto, A. Elemental and isotopic fractionation of type B calcium-aluminum-rich inclusions: Experiments, theoretical considerations, and constraints on their evolution. *Geochim. Cosmochim. Acta* **66**, 521–540 (2002).
11. Alexander, C. M. O'D. Chemical equilibrium and kinetic constraints for chondrule and CAI formation conditions. *Geochim. Cosmochim. Acta* **68**, 3943–3969 (2004).
12. Hewins, R. H., Connolly, H. C. Jr, Lofgren, G. E. & Libourel, G. in *Chondrules and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) Vol. 341, 286–316 (Astron. Soc. Pacific Conf. Ser., San Francisco, 2005).
13. Alexander, C. M. O'D. et al. The lack of potassium isotopic fractionation in Bishunpur chondrules. *Meteorit. Planet. Sci.* **35**, 859–868 (2000).
14. Nagahara, H. & Ozawa, K. Isotopic fractionation as a probe of heating processes in the solar nebula. *Chem. Geol.* **169**, 45–68 (2000).
15. Young, E. D. & Galy, A. The isotope chemistry and cosmochemistry of magnesium. *Rev. Min. Geochem.* **55**, 197–235 (2004).
16. Ebel, D. S. & Grossman, L. W. Condensation in dust-enriched systems. *Geochim. Cosmochim. Acta* **64**, 339–366 (2000).
17. Galy, A., Young, E. D., Ash, R. D. & O'Nions, R. K. The formation of chondrules at high gas pressures in the solar nebula. *Science* **290**, 1751–1754 (2000).
18. Young, E. D. Isotopic cosmobarometry—a synthesis of concepts and implications for chondrule and CAI formation mechanisms. *Lunar Planet. Sci. Conf. XXXV*, abstr. 1300 [CDROM] (2004).
19. Jones, R. H., Lee, T., Connolly, H. C. Jr, Love, S. G. & Shang, H. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) (Univ. Arizona Press, 2000).
20. Ciesla, F. J. & Hood, L. L. The nebular shock wave model for chondrule formation: shock processing in a particle-gas suspension. *Icarus* **158**, 281–293 (2002).
21. Cuzzi, J. N., Hogan, R. C., Paque, J. M. & Dobrovolskis, A. R. Size-selective concentration of chondrules and other small particles in protoplanetary nebula turbulence. *Astrophys. J.* **546**, 496–508 (2001).
22. Hood, L. L., Ciesla, F. J. & Weidenschilling, S. J. in *Chondrules and the Protoplanetary Disk* (eds Krot, A. N., Scott, E. R. D. & Reipurth, B.) Vol. 341, 873–882 (Astron. Soc. Pacific Conf. Ser., San Francisco, 2005).
23. Cuzzi, J. N. & Zahnle, K. R. Material enhancement in protoplanetary nebulae by particle drift through evaporation fronts. *Astrophys. J.* **614**, 490–496 (2004).
24. Ciesla, F. J. & Cuzzi, J. N. The evolution of the water distribution in a viscous protoplanetary disk. *Icarus* **181**, 178–204 (2006).
25. Yurimoto, H. & Kuramoto, N. Molecular cloud origin for the oxygen isotope heterogeneity in the solar system. *Science* **305**, 1763–1766 (2004).
26. Lyons, J. R. & Young, E. D. CO self-shielding as the origin of oxygen isotopic anomalies in the early solar nebula. *Nature* **435**, 317–320 (2005).
27. Krot, A. N. et al. Evolution of oxygen isotopic composition in the inner solar nebula. *Astrophys. J.* **622**, 1333–1342 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Young for several discussions and A. Boss, L. Nittler, and F. Ciesla for comments which improved the manuscript. This work was supported by NASA's Planetary Geology and Geophysics, Origins of Solar Systems, and Cosmochemistry programmes.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.N.C. (jcuzzi@mail.arc.nasa.gov).

LETTERS

Detection of magnetic circular dichroism using a transmission electron microscope

P. Schattschneider¹, S. Rubino¹, C. Hébert¹, J. Rusz², J. Kuneš², P. Novák², E. Carlino³, M. Fabrizioli^{3,4}, G. Panaccione³ & G. Rossi^{3,5}

A material is said to exhibit dichroism if its photon absorption spectrum depends on the polarization of the incident radiation. In the case of X-ray magnetic circular dichroism (XMCD), the absorption cross-section of a ferromagnet or a paramagnet in a magnetic field changes when the helicity of a circularly polarized photon is reversed relative to the magnetization direction. Although similarities between X-ray absorption and electron energy-loss spectroscopy in a transmission electron microscope (TEM) have long been recognized, it has been assumed that extending such equivalence to circular dichroism would require the electron beam in the TEM to be spin-polarized. Recently, it was argued on theoretical grounds that this assumption is probably wrong¹. Here we report the direct experimental detection of magnetic circular dichroism in a TEM. We compare our measurements of electron energy-loss magnetic chiral dichroism (EMCD) with XMCD spectra obtained from the same specimen that, together with theoretical calculations, show that chiral atomic transitions in a specimen are accessible with inelastic electron scattering under particular scattering conditions. This finding could have important consequences for the study of magnetism on the nanometre and subnanometre scales, as EMCD offers the potential for such spatial resolution down to the nanometre scale while providing depth information—in contrast to X-ray methods, which are mainly surface-sensitive.

X-ray methods have been extremely helpful to the improvement of understanding of magnetic phenomena in correlated electron systems. XMCD^{2–4} was first seen in the K-edge signal of Fe (ref. 5). Applications profited from the discovery that the dichroic signal in L_{2,3} edges was an order of magnitude higher than in K edges because of the strong spin–orbit interaction in the initial state⁶. X-ray magnetic dichroism has developed in the past 20 years to become one of the highest impact methods based on synchrotron radiation: it has made practical the concept of ‘local magnetometry’ by coupling the element specificity of X-ray absorption spectroscopy and magnetic moment analysis based on the sum-rules. Orbital, spin and quadrupolar moments can be retrieved notably from L_{2,3} edges of transition metals and from M_{4,5} edges of rare earths and actinides^{7–14}.

Microscopy applications of XMCD have been developed both by focusing soft X-rays by means of diffractive optics (Fresnel zone plates) or by passing the photoemitted electrons through electron-optical lenses to form high resolution images. In both cases lateral resolutions of about 100 nm can be achieved routinely, providing valuable information for mesoscaled systems. In a few cases 50 nm resolution has been demonstrated, with expectations of reaching 10 nm in the next few years. XMCD-photoemitted electron microscopy has a signal depth of a few nanometres only, and is thus

essentially a surface-sensitive technique. Other imaging techniques, such as electron holography in the TEM or Lorentz microscopy, have lateral resolution of <10 nm but are not element-specific.

The first step in studying linear magnetic dichroism in a TEM was taken in 1997¹⁵. This discovery was made possible by the earlier consideration¹⁶ that the polarization vector ϵ in X-ray absorption spectroscopy is formally equivalent to the momentum transfer $\hbar\mathbf{q}$ in electron scattering, where $\mathbf{q}$ is the wave vector transfer. A further step can be taken if we extend such equivalence to circularly polarized photons considered as a superposition of two linearly polarized waves dephased by 90° (formally represented by the imaginary unit i). We aim therefore at $\mathbf{q} \pm i\mathbf{q}'$ with $\mathbf{q}$ perpendicular to $\mathbf{q}'$ and $|\mathbf{q}| = |\mathbf{q}'|$, where $\hbar\mathbf{q}, \hbar\mathbf{q}'$ are momentum transfers in electron impact ionization. The inelastic electron scattering cross-section is proportional to a sum over squared matrix elements of the form $|\langle f|\mathbf{q} \cdot \mathbf{r}|i\rangle|^2$ for dipole allowed transitions from initial atomic states $|i\rangle$ to final empty states $|f\rangle$. Expanding $|\langle f|(\mathbf{q} + i\mathbf{q}') \cdot \mathbf{r}|i\rangle|^2$ we obtain two direct terms and one interference term. The situation is similar to the double slit experiment where a coupling term between plane waves with wave vectors $\mathbf{q}$ and $\mathbf{q}'$ gives rise to interference fringes. In this case the double differential inelastic scattering cross-section σ with respect to energy loss E and solid angle Ω for fast electrons reads:

$$\frac{\partial^2 \sigma}{\partial E \partial \Omega} \propto (1/q^4)[S(\mathbf{q}, E) + S(\mathbf{q}', E) + 2\text{Im}\{S(\mathbf{q}, \mathbf{q}', E)\}]$$

The first two terms in the square brackets are dynamic form factors. They describe inelastic scattering of an incident plane wave into an outgoing plane wave with wave vector transfer $\mathbf{q}$ and energy loss E . They appear in angle resolved electron energy-loss spectroscopy (EELS) and would produce directional dependence of spectra in anisotropic materials. The last term in the square brackets is the mixed dynamic form factor¹⁷. A more detailed explanation can be found in sections 2 and 3 of the Supplementary Information.

The interference term is ultimately responsible for the dichroism: the two incident plane electron waves are dephased by a quarter of a wavelength, causing the appearance of the imaginary part in the equation. This can be achieved by the set-up shown in Fig. 1. The energy loss spectra are measured in the diffraction plane with simultaneous wave vector transfer of $\mathbf{q}$ and $\mathbf{q}'$ from the coherent plane wave sources (000) and (110). The necessary phase shift of $2\pi\delta z/\lambda = \pi/2$ (where λ is wavelength and δz the difference in the optical path) between the transfer vectors is realized by an excitation error of $w = 0$, that is, in the exact two-beam case of a systematic row of Bragg spots. Shifting the detector from position ‘+’ to ‘–’ is equivalent to the exchange of $\mathbf{q}$ and $\mathbf{q}'$, thus reversing the sign of the phase shift. The situation is more complicated in practice, because the inelastic scattering event must be integrated over the thickness of

¹Service Centre for Transmission Electron Microscopy, Wiedner Hauptstraße 8-10/052, and Institut für Festkörperphysik, Wiedner Hauptstraße 8-10/138, Technische Universität Wien, A-1040 Wien, Austria. ²Institute of Physics ASCR, Cukrovarnická 10, 16253 Praha 6, Czech Republic. ³TASC INFN-CNR National Laboratory, Area Science Park, S.S.14, Km 163.5, I-34012 Trieste, Italy. ⁴Università degli Studi di Trieste, Piazzale Europa 1, I-34100 Trieste, Italy. ⁵Dipartimento di Fisica dell’Università di Modena e Reggio Emilia-I-41100 Modena, Italy.

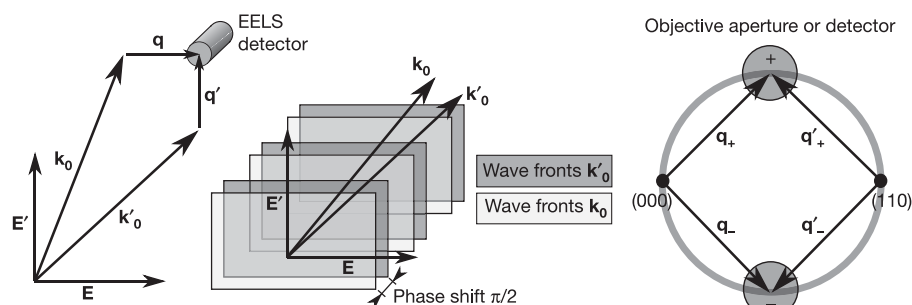


Figure 1 | Scattering geometry. Simplified scattering geometry, drawn in the diffraction plane of the TEM. Left: the detector (or a contrast aperture) selects q and q' . Centre: Bragg diffraction creates a coherent superposition of two incident plane waves (with wave vectors k_0 and k'_0 and electric field vectors E and E') in the Fe crystal. A phase shift of $\pi/2$ is set between the two wave fronts by tilting the incoming beam. Right: the larger circle

represents the points for which q is perpendicular to q' . The two smaller circles show the two positions for which the condition $|q_+| = |q_-|$ is also true and indicate the actual experimental set-up. As the two positions '+' and '-' have opposite chirality, EMCD can be detected by simply acquiring spectra at the two positions and taking their difference.

the specimen, thus smearing out the expected phase shift due to the *pendellösung* oscillations (sinusoidal intensity oscillations caused by the superposition of Bloch waves), and because the probe electron after scattering is also in a Bloch wave state. For thin specimens, however, the phase shift is $\pm\pi/2$ with good accuracy.

We chose Fe for the experiment because the Fe $L_{2,3}$ ionization edge shows a strong dichroic effect. In X-ray absorption spectroscopy it shows up as a variation of the intensity ratio of the L_3 and the L_2 edges when the helicity of the photon is changed. The magnitude of the dichroism is represented by the difference between the two spectra.

A (10 ± 2) -nm-thick Fe single crystal film was epitaxially grown in ultrahigh vacuum on top of a GaAs [001] self-supporting substrate previously thinned and ion milled to electron transparency and protected by 2.5 nm of Cu. Remanent in-plane magnetization could be evidenced by measurements of the transverse and longitudinal magneto-optic Kerr effect: the hysteresis loops indicated a coercive field of 80 Oe and full remanence in the (100) in-plane easy magnetization direction. The specimens, suitable for both X-ray absorption spectroscopy and TEM experiments, were transferred in ultrahigh vacuum to the APE-HE (High Energy) endstation for XMCD measures. Circularly polarized X-rays from the APPLE-II-type undulator radiation source at the ELETTRA storage ring were focused on a $50 \mu\text{m}$ spot on the sample surface, at 45° incidence. The dichroic signal was obtained by scanning in energy over the Fe $L_{2,3}$ edge and by reversing the photon helicity, as well as (for a given X-ray

helicity) by rotating the sample by 180° around an axis perpendicular to its surface. The remanent in-plane magnetization of the sample was mapped by XMCD all over the relevant part of the sample in order to assess the uniformity of its magnetization. Representative data are plotted in Fig. 2. A further capping layer of 2 nm of Cu was deposited to prevent Fe layer oxidation during the transferring of the samples to the TEM in Vienna.

EELS spectra were taken on a FEI Tecnai F20-FEGTEM S-Twin equipped with a Gatan image filter. A flat region of 100 nm radius and uniform thickness was selected in a single grain of Fe. Chemical microanalysis revealed negligible traces of contaminants (C, O and Mo). The sample is immersed in the magnetic field of the TEM objective lens pole piece, which is approximately 2 T and oriented perpendicular to the surface. The magnetization of the iron film in the TEM experiment is therefore saturated in the out-of-plane direction by a field that is large with respect to the in-plane coercivity. This is crystallographically identical to the in-plane magnetization used in the XMCD experiment, providing two physically equivalent conditions. The measured spectra are shown in Fig. 3. The dichroic signal is given by the difference between the two spectra taken at the position '+' and '-'.

For the simulations we used the electronic structure calculated with the WIEN2k code¹⁸—a full-potential, fully relativistic augmented plane wave code based on density functional theory. EMCD spectra require evaluation of the mixed dynamic form factor. For this

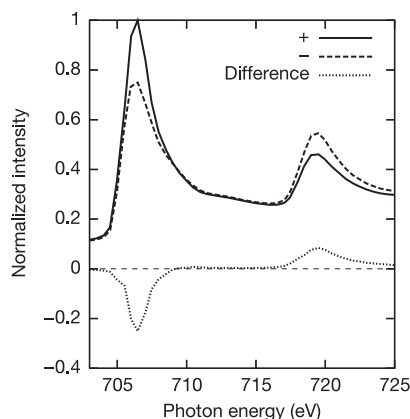


Figure 2 | X-ray magnetic circular dichroism. Circular dichroism in the Fe $L_{2,3}$ edge of epitaxial iron on GaAs(001) remanently magnetized along the in-plane [100] direction. The full and dashed curves are obtained by reversing the handedness (+/−) of circular polarized X-rays at each energy. The magnitude of the dichroism is represented by the difference (dotted) spectrum.

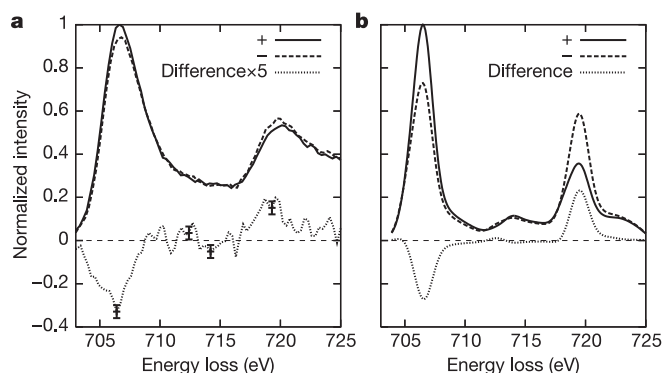


Figure 3 | Energy-loss magnetic chiral dichroism. Measured (a) and simulated (b) Fe $L_{2,3}$ edges for 10 nm Fe on GaAs (001) in the two configurations '+' and '-' explained in Fig. 1. The difference (magnified by a factor of 5 in the figure) is 0.07 for the measured spectra and 0.32 for the simulations. The r.m.s. of the noise, indicated by the error bars in a, is $\pm 8\%$ of the difference.

purpose, we extended the computer code for the calculation of inelastic interference terms^{19,20} to the case of two perpendicular momentum transfers $\hbar\mathbf{q}$ and $\hbar\mathbf{q}'$. Details about the calculation can be found in section 5 of the Supplementary Information.

A comparison (Fig. 3) with the TEM experiment shows that the observed dichroic signal is smaller than predicted. This can be qualitatively understood by considering the effects of the simplifications we used: as the $\mathbf{q}$ -selecting aperture is placed on the (200) spot of the GaAs substrate, a large part of the collected signal is non-dichroic Bragg and thermal diffuse scattering. The (200) GaAs spot contributes more than 50% of non-dichroic intensity by coupling of quasi-elastic (Bragg) to inelastic scattering. Moreover, the integration of the signal over the finite spectrometer entrance aperture and the non-zero convergence angle reduces the dichroic effect. Another prominent effect is the sensitivity of the signal to the specimen thickness. A thickness variation of ± 1 nm within the sampling area changes the dichroic signal by another 20%, as shown in Supplementary Fig. 3.

Clearly both EMCD in the TEM and XMCD on the Fe/GaAs sample show comparable spectroscopic features, caused by the respective forced or remanent magnetization. Furthermore, the comparison with the *ab initio* calculation gives the first experimental confirmation of the EMCD effect in Fe. Provided that the dichroic signal can be increased by optimizing the scattering geometry, EMCD could evolve into a new method for the analysis and characterization of magnetic materials at the nanometre scale. It might unravel the complexity of phase separation in highly correlated electron systems, and improve our understanding of certain magnetic phenomena in the manganites²¹, dilute magnetic semiconductors, metamagnetic materials and ferromagnetic/antiferromagnetic interfaces in magnetic multilayers. In combination with the analytical and crystallographic tools available in a TEM, EMCD may become a key microscopic method in spintronics and nanomagnetism.

Received 13 September 2005; accepted 29 March 2006.

- Hébert, C. & Schattschneider, P. A proposal for dichroic experiments in the electron microscope. *Ultramicroscopy* **96**, 463–468 (2003).
- Thole, B. P. *et al.* X-ray circular dichroism as a probe of orbital magnetization. *Phys. Rev. Lett.* **68**, 1943–1946 (1992).
- Carra, P. *et al.* X-ray circular dichroism and local magnetic fields. *Phys. Rev. Lett.* **70**, 694–697 (1993).
- van der Laan, G. *et al.* Experimental proof of magnetic X-ray dichroism. *Phys. Rev. B* **34**, 6529–6531 (1986).
- Schütz, G. *et al.* Absorption of circularly polarized X-rays in iron. *Phys. Rev. Lett.* **58**, 737–740 (1987).
- Chen, C. T. *et al.* Experimental confirmation of the X-ray magnetic circular dichroism sum rules for iron and cobalt. *Phys. Rev. Lett.* **75**, 152–155 (1995).
- Ohldag, H. *et al.* Spectroscopic identification and direct imaging of interfacial magnetic spins. *Phys. Rev. Lett.* **87**, 247201–247204 (2001).
- Wu, Y. Z. *et al.* Enhanced orbital magnetic moment on Co atoms in Co/Pd multilayers: a magnetic circular X-ray dichroism study. *Phys. Rev. Lett.* **69**, 2307–2310 (1992).
- Tischer, M. *et al.* Enhancement of orbital magnetism at surfaces: Co on Cu(100). *Phys. Rev. Lett.* **75**, 1602–1605 (1995).
- Won, C. *et al.* Magnetic phase transition in Co/Cu/Ni/Cu(100) and Co/Fe/Ni/Cu(100). *Phys. Rev. Lett.* **91**, 147202–147205 (1993).
- Weller, D. *et al.* Microscopic origin of magnetic anisotropy in Au/Co/Au probed with X-ray magnetic circular dichroism. *Phys. Rev. Lett.* **75**, 3752–3755 (1995).
- Vogel, J. & Sacchi, M. Magnetic moments in as-deposited and annealed Ni layers on Fe(001): an x-ray-dichroism study. *Phys. Rev. B* **53**, 3409–3414 (1996).
- Stöhr, J. Exploring the microscopic origin of magnetic anisotropies with X-ray magnetic circular dichroism (XMCD) spectroscopy. *J. Magn. Magn. Mater.* **200**, 470–497 (1999).
- Stöhr, J. X-ray magnetic circular dichroism spectroscopy of transition metal thin films. *J. Electron Spectrosc. Related Phenom.* **75**, 253–272 (1995).
- Yuan, J. & Menon, N. K. Magnetic linear dichroism in electron energy loss spectroscopy. *J. Appl. Phys.* **81**, 5087–5089 (1997).
- Hitchcock, A. P. Near edge electron energy loss spectroscopy: comparison to X-ray absorption. *Jpn. J. Appl. Phys.* **32** (suppl. 2), 176–181 (1992).
- Kohl, H. & Rose, H. Theory of image formation by inelastically scattered electrons in the electron microscope. *Adv. Electron. Electron Phys.* **65**, 173–227 (1985).
- Blaha, P., Schwarz, K., Madsen, G. K. H., Kvasnicka, D. & Luitz, J. *An Augmented Plane Wave and Local Orbitals Program for Calculating Crystal Properties* (ed. Schwarz, K.) (Technical University of Wien, Vienna, 2001).
- Nelhiebel, M., Schattschneider, P. & Jouffrey, B. Observation of ionization in a crystal interferometer. *Phys. Rev. Lett.* **85**, 1847–1850 (2000).
- Nelhiebel, M. *et al.* Theory of orientation sensitive near-edge fine structure core-level spectroscopy. *Phys. Rev. B* **59**, 10959–10969 (1999).
- Mathur, N. & Littlewood, P. Mesoscopic texture in manganites. *Phys. Today* **56**, 25–30 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Bauer and S. Buehler-Paschen for discussions. This work was supported by the European Union under the project CHIRALTEM.

Author Contributions P.S. and C.H. had the idea to do this experiment and developed the principles; P.S. wrote the manuscript; S.R. performed the experiment on the TEM and adapted the EMCD program; J.R. was the principal developer of the EMCD program; J.K. helped substantially in debugging the EMCD program; P.N. generalized two programs contained in the WIEN2k package that were used in the EMCD program and helped in its debugging and implementation; E.C. produced and characterized the TEM substrates; M.F. grew the iron films and ran XMCD and magneto-optic Kerr effect experiments; G.P. and G.R. coordinated the work at the synchrotron and the XMCD data analysis. All authors discussed the results and contributed to the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.S. (schatt@ifp.tuwien.ac.at).

Ge/Si nanowire heterostructures as high-performance field-effect transistors

Jie Xiang^{1*}, Wei Lu^{1*}, Yongjie Hu¹, Yue Wu¹, Hao Yan¹ & Charles M. Lieber^{1,2}

Semiconducting carbon nanotubes^{1,2} and nanowires³ are potential alternatives to planar metal-oxide-semiconductor field-effect transistors (MOSFETs)⁴ owing, for example, to their unique electronic structure and reduced carrier scattering caused by one-dimensional quantum confinement effects^{1,5}. Studies have demonstrated long carrier mean free paths at room temperature in both carbon nanotubes^{1,6} and Ge/Si core/shell nanowires⁷. In the case of carbon nanotube FETs, devices have been fabricated that work close to the ballistic limit⁸. Applications of high-performance carbon nanotube FETs have been hindered, however, by difficulties in producing uniform semiconducting nanotubes, a factor not limiting nanowires, which have been prepared with reproducible electronic properties in high yield as required for large-scale integrated systems^{3,9,10}. Yet whether nanowire field-effect transistors (NWFETs) can indeed outperform their planar counterparts is still unclear⁴. Here we report studies on Ge/Si core/shell nanowire heterostructures configured as FETs using high- κ dielectrics in a top-gate geometry. The clean one-dimensional hole-gas in the Ge/Si nanowire heterostructures⁷ and enhanced gate coupling with high- κ dielectrics give high-performance FETs values of the scaled transconductance ($3.3 \text{ mS } \mu\text{m}^{-1}$) and on-current ($2.1 \text{ mA } \mu\text{m}^{-1}$) that are three to four times greater than state-of-the-art MOSFETs and are the highest obtained on NWFETs. Furthermore, comparison of the intrinsic switching delay, $\tau = CV/I$, which represents a key metric for device applications^{4,11}, shows that the performance of Ge/Si NWFETs is comparable to similar length carbon nanotube FETs and substantially exceeds the length-dependent scaling of planar silicon MOSFETs.

Silicon^{9,10,12} and germanium^{13,14} nanowires have been the focus of recent studies of one-dimensional (1D) FETs. However, metal contacts to single-component nanowires generally produce Schottky barriers that limit device performance¹⁵, and moreover, scattering from charged dopants can also reduce the intrinsic mobility of these nanowire devices¹⁵. In contrast, we have recently demonstrated transparent contacts and low-bias ballistic transport⁷ in undoped Ge/Si core/shell nanowire heterostructures (Fig. 1a, b), with an estimated scattering mean free path of $\sim 500 \text{ nm}$. The 1D sub-band spacing in the typical 15-nm core Ge/Si nanowires determined through both experimental measurements and theoretical calculations⁷ is $\sim 25 \text{ meV}$, and thus at room temperature several sub-bands may participate in NWFET transport. While the Ge/Si nanowire devices will not be strictly 1D, the limited number of conduction channels and clean material structure can benefit performance through, for example, a reduction in scattering. To explore the potential of Ge/Si nanowire heterostructures as high-performance FETs we have fabricated (see Methods) devices using thin HfO_2 or ZrO_2 high- κ gate dielectrics and metal top gate electrodes (Fig. 1c, d). Cross-sectional transmission electron microscopy (TEM) images (Fig. 1e) show that both the

high- κ and metal top gate conform to the approximately circular cross-section of the nanowire, and also verify the Ge/Si core/shell structure. The conformal top gate structure approaches an ideal cylindrical gate geometry, and together with the high- κ dielectrics produces a much more efficient gate response than previous studies using lower- κ SiO_2 dielectric and planar back gates^{9,12,16}.

Typical output and transfer characteristics recorded from a Ge/Si device fabricated in this way with a channel length, $L = 1 \mu\text{m}$ and a total diameter of 18 nm (device A) are shown in Fig. 2a, b. The family of I_d - V_{ds} curves (Fig. 2a) show that the drain current I_d first increases then saturates with increasingly negative drain voltage, similar to a conventional long channel MOSFET¹¹. These data also show that I_d increases as the gate voltage V_g decreases from 1 to -2 V , and thus that the device is a p -type depletion-mode FET. This p -type FET behaviour is expected from the band diagram in Fig. 1b, where the Fermi level lies below the Ge valence band edge in the absence of a gate. The I_d - V_g transfer curve recorded for the drain bias voltage

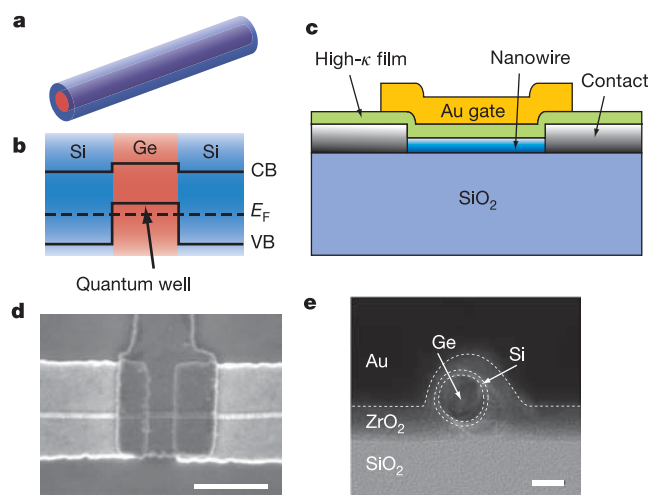


Figure 1 | Ge/Si core/shell NWFET. **a**, Schematic of a Ge/Si core/shell nanowire. **b**, Cross-sectional diagram showing the formation of hole-gas in the Ge quantum well confined by the epitaxial Si shell, where CB is the conduction band and VB is the valence band. The dashed line indicates the Fermi level, E_F . The valence band offset of $\sim 500 \text{ meV}$ between Ge and Si serves as a confinement potential to the hole-gas as discussed previously⁷. **c**, Schematic of the NWFET device with high- κ dielectric layer and Au top gate. **d**, Top-view SEM image of a typical device. The Au top gate overlaps with the Ni source/drain electrodes to ensure full coverage of the channel. Scale bar, 500 nm. **e**, Cross-sectional TEM image of a device prepared using 7 nm ZrO_2 dielectric. Dotted lines are guides to the eye showing boundaries between different materials denoted in the image. The nanowire is tilted off the imaging axis. Scale bar, 10 nm.

¹Department of Chemistry and Chemical Biology, ²Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA.

*These authors contributed equally to this work.

$V_{ds} = -1$ V (Fig. 2b) demonstrates that the NWFET has a peak transconductance, $g_m = dI_d/dV_g$, of $26 \mu\text{S}$. In addition, the device exhibits a maximum drain current $I_{d(\text{max})}$ of $35 \mu\text{A}$ at $V_g = -2$ V. We note these values of g_m and $I_{d(\text{max})}$ substantially exceed the

best performance reported to date in single semiconductor NWFETs^{12,14}.

The on current I_{on} for a FET device is usually determined at $V_g = V_{ds} = V_{dd}$, where V_{dd} is the power supply voltage and equals 1 V in our case. Following conventions in planar devices, we define on and off currents as the values measured at $V_{g(\text{on})} = V_T - 0.7V_{dd}$ and $V_{g(\text{off})} = V_T + 0.3V_{dd}$, so that 30% of the V_g swing above the threshold voltage V_T is applied to turn the device off, while the remaining 70% sets the operation range of the on state (Fig. 2b). Similar methods have been proposed in benchmarking carbon nanotube FET devices^{4,17}. From Fig. 2b, we obtain $I_{on} = 14 \mu\text{A}$ for this $1\text{-}\mu\text{m}$ -long device. Significantly, the scaled values of g_m and I_{on} , $1.4 \text{ mS } \mu\text{m}^{-1}$ and $0.78 \text{ mA } \mu\text{m}^{-1}$, using the total nanowire diameter as the device width, already exceeds the values of $0.8 \text{ mS } \mu\text{m}^{-1}$ and $0.71 \text{ mA } \mu\text{m}^{-1}$ recently reported in much shorter, sub-100-nm silicon *p*-MOSFETs employing high- κ dielectrics¹⁸.

In addition, we have prepared and studied a large number of Ge/Si NWFET devices with L varying from $1 \mu\text{m}$ to 190 nm ; essentially all of these devices exhibited high-performance behaviour and testify to the reproducibility of both the Ge/Si nanowires and contacts to this material. Representative data obtained from a $L = 190 \text{ nm}$ device (device B), which should exhibit larger g_m and I_d values owing to reduced channel resistance, are shown in Fig. 2c. These data yield $g_m = 60 \mu\text{S}$, $I_{on} = 37 \mu\text{A}$ ($V_{dd} = 1$ V), and $I_{d(\text{max})} = 91 \mu\text{A}$, and correspond to scaled values of g_m and I_{on} of $3.3 \text{ mS } \mu\text{m}^{-1}$ and $2.1 \text{ mA } \mu\text{m}^{-1}$, respectively. Notably, these values are more than twice that achieved in the longer channel device and are 3–4 times greater than state-of-the-art Si *p*-MOSFETs¹⁸. The geometric gate capacitance per unit area in our NWFETs, $44 \text{ fF } \mu\text{m}^{-2}$ (Methods), is only 29% larger than the $34 \text{ fF } \mu\text{m}^{-2}$ in these Si *p*-MOSFETs¹⁸. Therefore the large gain in g_m and I_{on} cannot be accounted for by an increase in gate capacitance alone. Moreover, the hole mobility for this Ge/Si NWFET, $730 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, extracted at the linear region ($|V_{ds}| = 10 \text{ mV}$) from the peak $g_m = 3 \mu\text{S}$ at $|V_g - V_T| = 0.13 \text{ V}$ using the charge control model, represents an improvement of more than a factor of ten over that of the Si *p*-MOSFET with HfO_2 gate dielectric ($50\text{--}60 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$)¹⁸, and also is more than twice the reported low-field mobility of Ge and strained SiGe heterostructure PMOS devices^{19,20}. Improved mobility is observed for NWFETs with channel lengths from 0.19 to $1 \mu\text{m}$ (Supplementary Fig. S1), with an average of $640 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. These improvements over planar device structures thus verify the performance benefit due to the quasi-1D transport in clean Ge/Si heterostructure nanowires.

The subthreshold region of the I_d - V_g data was also analysed and yields values of the slope, $S = -[d(\log_{10} I_d)/dV_g]^{-1}$, of 105 and 100 mV per decade for the $L = 1 \mu\text{m}$ and 190 nm NWFETs, respectively, for $V_{ds} = -1$ V (Fig. 2b, c). Similar values of S were obtained from I_d - V_g data recorded on both devices using V_{ds} from -0.01 to -1 V, which indicate the absence of significant short-channel effects¹¹ for devices down to at least $L = 190 \text{ nm}$ and excellent V_g control of the channel potential over the competing effect of drain-induced barrier lowering at larger biases¹¹.

In general, an FET with a small S is essential for modern logic circuits as it reduces the off state current and minimizes static power dissipation. The value of S can be estimated¹¹ by $2.3k_B T/e\alpha$, where T is temperature and α is the gate-coupling factor, which yields a room temperature minimum ($\alpha = 1$) of 60 mV per decade. The values of S determined for the $L = 1 \mu\text{m}$ and 190 nm Ge/Si NWFETs are superior to the best value (140 mV per decade) reported previously⁹ for NWFETs but still larger than the theoretical minimum. The non-ideal gate coupling ($\alpha < 1$), which yields this larger S value, is probably due to a finite trap state density at the nanowire/high- κ interface²¹. Optimization of the high- κ deposition process during fabrication or growth of a cylindrical high- κ shell on the Ge/Si nanowire before fabrication should yield improved interface quality and enable us to approach ideal subthreshold behaviour in the future in these NWFETs.

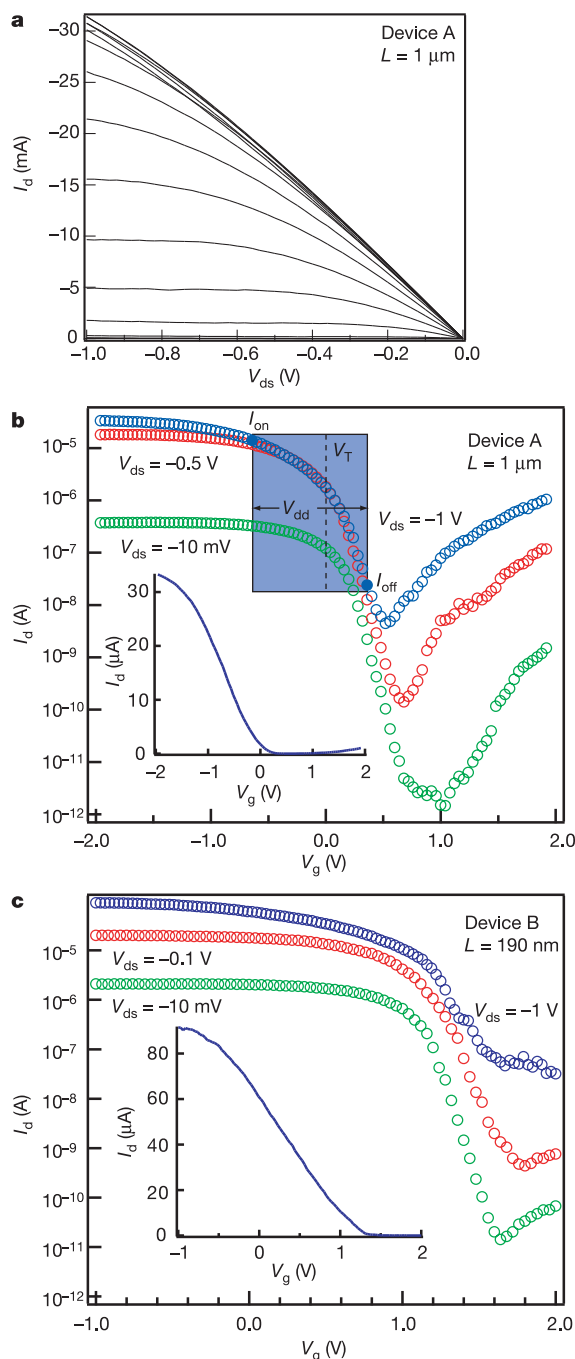


Figure 2 | Characteristics of high-performance Ge/Si NWFET. **a**, I_d - V_{ds} data for device A ($L = 1 \mu\text{m}$, 4 nm HfO_2 dielectric) with $V_g = -2$ to 2 V in 0.25 V steps from top to bottom. **b**, I_d - V_g for device A with blue, red, and green data points corresponding to V_{ds} values of -1 , -0.5 and -0.01 V, respectively. The leakage current through the gate electrode (I_g) is $< 10^{-10} \text{ A}$, which excludes I_g as source of increase in I_d at $V_g > \sim 0.5$ V. Inset, linear scale plot of I_d versus V_g measured at $V_{ds} = -1$ V. The blue-shaded area defines the 1 V gate voltage window described in the text, where V_T was determined from the intercept of the tangent of maximum slope (linear transconductance) region of the I_d - V_g curve¹¹. **c**, I_d - V_g data for device B ($L = 190 \text{ nm}$, 4 nm HfO_2 dielectric) with blue, red and green data points corresponding to V_{ds} values of -1 , -0.1 and -0.01 V, respectively. Inset, linear scale plot of I_d versus V_g measured at $V_{ds} = -1$ V.

An important benchmark of transistor performance is the intrinsic delay, $\tau = CV/I$, where C is the gate capacitance, $V = V_{dd}$, and I is on current I_{on} . As defined, τ represents the fundamental RC (where R is the device resistance and C is the capacitance) delay of the device and provides a frequency limit for transistor operation that is relatively insensitive to gate dielectrics and device width, and thus represents a good parameter for comparing different types of devices¹¹. The calculated intrinsic delays are 57 and 4 ps for devices A and B in Fig. 2, respectively, where C was determined by numerical simulation (Methods). A summary of the results from seven Ge/Si NWFETs versus L and the corresponding scaling for Si MOSFETs (Fig. 3a) highlights several key points. First, the data show clear speed advantage at a given L for the Ge/Si NWFETs versus Si p -MOSFETs. For example, the intrinsic delay for a 190 nm Si planar device is larger than 10 ps, about three times longer than our device B. Second, the delay time for the 190 nm Ge/Si device is about the same as that of similar-length CNTFET devices⁴. Last, length scaling of τ is more favourable for our Ge/Si NWFETs than Si MOSFETs (that is, slope of ~ 1.5 versus ~ 1.1). We attribute this important difference to a suppression of scattering in the quasi-1D quantum confined Ge/Si nanowires versus MOSFETs²², although additional studies will be needed to support this idea.

To capture the off state leakage current property, we used a method described by ref. 17 and studied the CV/I_{on} versus I_{on}/I_{off} ratio. Here we assume full control of the threshold voltage, allowing a window of $V_{g(on)} - V_{g(off)} = V_{dd} = 1$ V to move along the V_g axis and define a pair of I_{on} and I_{off} value from the $I_d - V_g$ data plot. The CV/I_{on} versus I_{on}/I_{off} data for devices A and B (Fig. 3b) shows the trade-off between high speed and small leakage. The smallest τ is observed at the largest I_{on} , although this corresponds to a minimum on/off ratio. As the on/off ratio increases τ also increases⁴, until the on/off ratio reaches a maximum limited by ambipolar conduction (see below). The arrows correspond to the intrinsic delay values obtained from the 70–30% criteria used to define I_{on} in the benchmark plot in Fig. 3a, and show that the I_{on}/I_{off} ratios for the A and B devices are 100 and 580, respectively. On/off ratios for the rest of the devices in Fig. 3a all lie within this range. The on/off ratio is expected to reach 10^4 – 10^5 as the subthreshold slope is improved to the ideal value of 60 mV per decade. Studies of strained SiGe planar devices show that subthreshold slopes of 66–70 mV per decade are achievable²⁰,

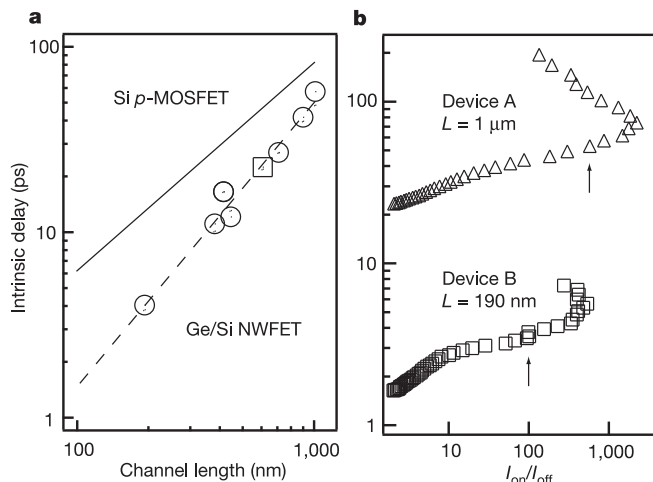


Figure 3 | Benchmark and comparison of Ge/Si FETs. **a**, Intrinsic delay τ versus channel length for seven different Ge/Si nanowire devices with HfO_2 dielectric (open circle) and ZrO_2 dielectric (open square). Data for devices A and B are included. The I_{on} values were measured at $V_{g(on)} = V_T - 0.7V_{dd}$ as discussed in the text. The dashed line is a fit to the data points while solid line is the Si p -MOSFET results from ref. 4. **b**, Intrinsic delay versus on/off ratio for the two devices in Fig. 2. Arrows indicate the values of intrinsic delay used in **a**.

although we note that the on/off ratio of 10^2 may already meet a lower practical limit for certain high-performance applications²³.

The above Ge/Si NWFETs are depletion-mode devices with threshold voltages $V_T > 0$, and require $V_g > V_T$ to be turned off. However, enhancement-mode FETs with $V_T < 0$, which are off for $V_g = 0$, are technologically more desirable because they consume less static power. In addition and as discussed above, obtaining the optimal device operation depends critically on the full control of the threshold voltage. We have exploited the top gated structure (Fig. 1c) to tune V_T through variations in the gate metal work function. Comparison of $I_d - V_g$ data recorded using Au and Al metal gates (Fig. 4a) clearly shows a change from depletion mode, $V_T = +0.65$ V, to enhancement mode $V_T = -0.65$ V, while other key device parameters remain the same. Measurements made on 68 NWFETs yield average threshold values of 0.53 ± 0.17 and

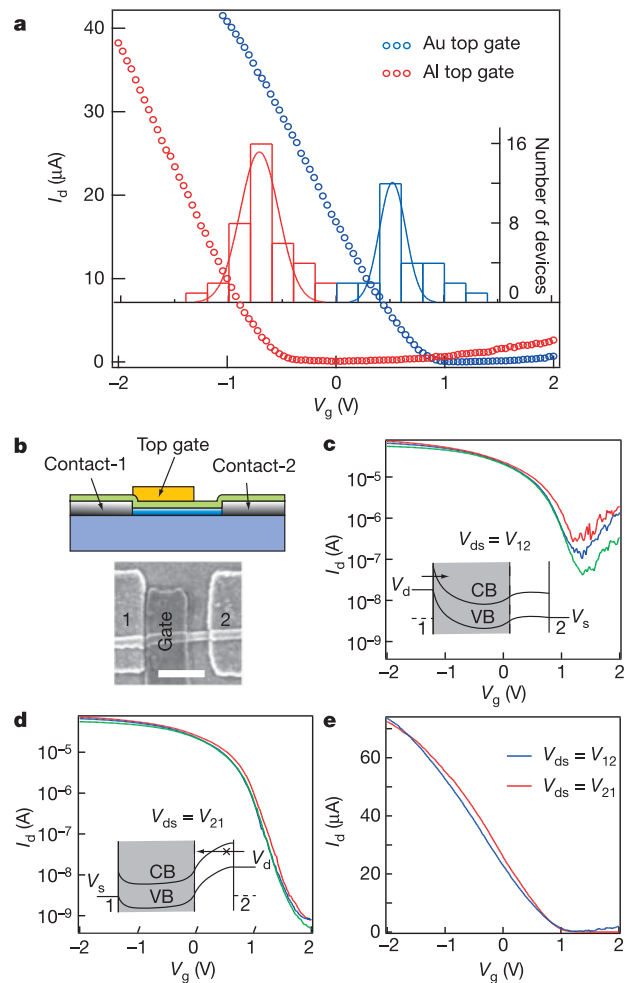


Figure 4 | Control of threshold voltage and ambipolar conduction through device design. **a**, $I_d - V_g$ curves for two $L = 300$ nm devices with Au (blue) and Al (red) top gate electrodes ($V_{ds} = -1$ V). Inset shows histogram of V_T with the same V_g axis for a total of 68 $L = 300$ nm devices with Au (blue) and Al (red) top gates. Solid lines correspond to gaussian fits to the two distributions. **b**, Schematic and SEM image of the asymmetrical gate structure designed to suppress ambipolar conduction. Scale bar, 300 nm. **c**, $I_d - V_g$ of partially gated device with ambipolar conduction; bias was applied to contact 1 ($V_{ds} = V_{12}$). Inset, schematic of band bending in the NWFET at finite bias. Arrow denotes electron injection at the drain contact. **d**, $I_d - V_g$ for $V_{ds} = V_{21}$. Inset, schematic of band bending with electron injection denoted by arrow. The red, blue and green curves in **c** and **d** correspond to V_{ds} values of -1 , -0.8 and -0.6 V, respectively. **e**, Linear scale $I_d - V_g$ ($V_{ds} = -1$ V) for the devices in **c** and **d**. The two devices have the same peak $g_m = 35 \mu\text{S}$ and $I_{d(max)} = 73 \mu\text{A}$.

-0.72 ± 0.25 V for Au and Al top gates, respectively, and thus demonstrate the reproducibility of this effect in our high-performance Ge/Si NWFETs. The V_T shift of 1.25 V corresponds approximately to the work function difference between Au ($\Phi_{\text{Au}} = 5.31$ – 5.47 eV) and Al ($\Phi_{\text{Al}} = 4.28$ eV) with small deviations attributable to metal/dielectric interface states²⁴. More generally, these results indicate that it will be possible to tune V_T for specific applications simply through a choice of top gate metal with specific work function in fabrication.

The Ge/Si NWFETs also exhibit an increase in I_d when V_g is increased to larger positive values (for example, Fig. 2) owing to conduction by electron carriers (versus holes). Similar ambipolar conduction has been observed in high-performance CNTFETs with metal contacts^{4,25,26}, and is deleterious for applications since it reduces the window of operation and increases the minimum off-state current. To address this issue we characterized devices with asymmetrical partial gates (Fig. 4b). Data recorded from a NWFET with bias voltage applied to contact 1 (proximal to the gate) and holding contact 2 at ground (Fig. 4c), $V_{\text{ds}} = V_{12}$, show ambipolar conduction like the fully gated device in Fig. 2. Significantly, switching the source and drain electrodes ($V_{\text{ds}} = V_{21}$) dramatically suppresses the ambipolar current from 300 to 0.8 nA at $V_{\text{ds}} = -1$ V (Fig. 4d). These results can be explained by the corresponding band diagrams (insets, Fig. 4c, d). In the first case, electron injection at the drain increases with increasing V_g and ultimately dominates the current, while in the second, the ungated region near contact 2 acts as a thick barrier to electron transport and suppresses electron current even at large downward bending of the conduction band.

Importantly, the reduction in ambipolar current using this device structure does not limit other key NWFET characteristics. The on state conductance and transconductance (Fig. 4e) show no degradation compared to fully gated devices with similar dimensions (for example, Fig. 4a), and S (Fig. 4d) shows little V_{ds} dependence, indicating excellent gate control¹¹. These observations contrast with experiments on CNTFETs with similar gate structures²⁷, which may have been limited by the presence of Schottky barriers at the CNT contacts. Such limitations do not exist for Ge/Si NWFETs, which do not have contact barriers⁷, and thus the asymmetrical gate structure can yield unipolar NWFETs without sacrificing performance.

In summary, we have demonstrated top-gated Ge/Si NWFET heterostructures with high- κ dielectrics that exhibit scaled transconductance and on-current values of $3.3 \text{ mS } \mu\text{m}^{-1}$ and $2.1 \text{ mA } \mu\text{m}^{-1}$, respectively, are three to four times greater than those for state-of-the-art MOSFETs. In addition, the Ge/Si NWFET hole mobility, $730 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, is more than a factor of ten greater than the Si p -MOSFET with HfO_2 gate dielectric¹⁸ and more than twice that of Ge and strained SiGe heterostructure PMOS devices^{19,20}. These values, together with the demonstrated control over threshold voltage and ambipolar behaviour, suggest substantial promise of Ge/Si NWFETs, although further performance improvements at the single device level should also be possible through optimization of gate coupling and size scaling²⁸. Looking to the future, the ability to prepare high-performance Ge/Si NWFETs in close to 100% yield, which represents a distinct advantage over similar performance CNTs, and assemble nanowires *en masse* in addressable arrays⁹ could open up advances and applications in several areas, including high-frequency electronics on plastic and glass substrates¹⁰, higher-sensitivity nanosensors²⁹, and possibly extending the roadmap for high-performance logic.

METHODS

Fabrication and measurement of Ge/Si NWFET devices. The growth of epitaxial core/shell Ge/Si nanowires and fabrication of Ni-contacted NWFETs are described elsewhere⁷. Nanowires have an average core diameter of 14.6 nm and Si shell thickness of 1.7 nm, and normally exhibit $\langle 110 \rangle$ growth direction. A thin high- κ dielectric film was deposited on the devices using the atomic layer

deposition (ALD) process. 30 cycles for HfO_2 deposition and 50 cycles for ZrO_2 were used at 110°C with each cycle consisting of 1 s water vapour pulse, 5 s N_2 purge, 3 s precursor, and 5 s N_2 purge. Tetrakis(dimethylamino) hafnium [$\text{Hf}(\text{N}(\text{CH}_3)_2)_4$], and tetrakis(dimethylamino) zirconium were used as precursors. Electron beam lithography was used to define the top gate, followed by thermal evaporation of either Cr/Au (5 nm/50 nm) or Al (50 nm). The devices were measured at room temperature in vacuum ($P < 10^{-4}$ torr) with a probe station (TTP-4, Desert Cryogenics).

Cross-sectional TEM sample preparation. Dry-transfer from the growth substrate was used to deposit aligned nanowire arrays with inter-nanowire spacings of several micrometres on a Si/SiO₂ wafer. The wafer was then coated with a thin film of ZrO_2 high- κ dielectric and Au metal as described above. Cross-sectional TEM samples were prepared by cutting the wafer into thin slices, followed by mechanical polishing and further thinning by ion milling. TEM images were taken by a JEOL 2010F high-resolution microscope.

Calculation of mobility and intrinsic delay CV/I . The gate capacitance, C , was calculated using numerical simulations on nanowire devices with a Ge core diameter of 14.6 nm and a Si shell thickness of 1.7 nm; these parameters were determined for devices using cross-sectional TEM measurements. The thickness for HfO_2 ($\kappa = 23$) and ZrO_2 ($\kappa = 20$) are 4 and 7 nm, respectively. Assuming the top gate conformally covers the top half of the nanowire as indicated by the cross-sectional TEM image, we obtained the gate capacitances per unit length of $C_L = 800 \text{ aF } \mu\text{m}^{-1}$ (HfO_2) and $580 \text{ aF } \mu\text{m}^{-1}$ (ZrO_2) from two-dimensional electrostatic simulations (Quickfield, Tera Analysis, Denmark). When scaled using the total diameter of the nanowire, we obtained gate capacitances per unit area of 44 and $32 \text{ fF } \mu\text{m}^{-2}$ for the HfO_2 and ZrO_2 dielectrics used, respectively. We note that the calculation tends to overestimate the gate coupling capacitance because it does not include the effect of quantum capacitance from the finite density of states in the 1D Ge channel³⁰, and does not consider the formation of interfacial silicon oxide layer that tends to reduce the κ value. Mobility is calculated from low-bias g_m based on the charge control model: $\mu = \frac{g_m}{V_{\text{ds}}} \cdot \frac{L}{C}$, where L is device gate length. Supplementary Fig. S1 shows a linear relationship between the inverse transconductance and the channel length for three different devices, consistent with this model. For the intrinsic delay CV/I_{on} , $V = V_{\text{dd}} = 1$ V is the power supply voltage for both the V_g swing and saturation bias.

Received 11 October 2005; accepted 10 April 2006.

- McEuen, P. L., Fuhrer, M. S. & Park, H. K. Single-walled carbon nanotube electronics. *IEEE Trans. Nanotechnol.* **1**, 78–85 (2002).
- Avouris, P. Molecular electronics with carbon nanotubes. *Acc. Chem. Res.* **35**, 1026–1034 (2002).
- Lieber, C. M. Nanoscale science and technology: Building a big future from small things. *MRS Bull.* **28**, 486–491 (2003).
- Chau, R. et al. Benchmarking nanotechnology for high-performance and low-power logic transistor applications. *IEEE Trans. Nanotechnol.* **4**, 153–158 (2005).
- McEuen, P. L., Bockrath, M., Cobden, D. H., Yoon, Y. G. & Louie, S. G. Disorder, pseudospins, and backscattering in carbon nanotubes. *Phys. Rev. Lett.* **83**, 5098–5101 (1999).
- Javey, A. et al. High-field quasiballistic transport in short carbon nanotubes. *Phys. Rev. Lett.* **92**, 106804 (2004).
- Lu, W., Xiang, J., Timko, B. P., Wu, Y. & Lieber, C. M. One-dimensional hole gas in germanium/silicon nanowire heterostructures. *Proc. Natl. Acad. Sci. USA* **102**, 10046–10051 (2005).
- Javey, A., Guo, J., Wang, Q., Lundstrom, M. & Dai, H. J. Ballistic carbon nanotube field-effect transistors. *Nature* **424**, 654–657 (2003).
- Jin, S. et al. Scalable interconnection and integration of nanowire devices without registration. *Nano Lett.* **4**, 915–919 (2004).
- Friedman, R. S., McAlpine, M. C., Ricketts, D. S., Ham, D. & Lieber, C. M. High-speed integrated nanowire circuits. *Nature* **434**, 1085 (2005).
- Taur, Y. & Ning, T. H. *Fundamentals of Modern VLSI Devices* (Cambridge Univ. Press, Cambridge, UK, 1998).
- Cui, Y., Zhong, Z. H., Wang, D. L., Wang, W. U. & Lieber, C. M. High performance silicon nanowire field effect transistors. *Nano Lett.* **3**, 149–152 (2003).
- Greytak, A. B., Lauhon, L. J., Gudiksen, M. S. & Lieber, C. M. Growth and transport properties of complementary germanium nanowire field-effect transistors. *Appl. Phys. Lett.* **84**, 4176–4178 (2004).
- Wang, D. W. et al. Germanium nanowire field-effect transistors with SiO₂ and high- κ HfO_2 gate dielectrics. *Appl. Phys. Lett.* **83**, 2432–2434 (2003).
- Zheng, G. F., Lu, W., Jin, S. & Lieber, C. M. Synthesis and fabrication of high-performance n-type silicon nanowire transistors. *Adv. Mater.* **16**, 1890–1893 (2004).
- Javey, A. et al. High- κ dielectrics for advanced carbon-nanotube transistors and logic gates. *Nature Mater.* **1**, 241–246 (2002).
- Guo, J., Javey, A., Dai, H. J. & Lundstrom, M. Performance analysis and design optimization of near ballistic carbon nanotube field-effect transistors. *IEDM Tech. Dig.*, 703–706 (2004).

18. Chau, R. *et al.* High- κ /metal-gate stack and its MOSFET characteristics. *IEEE Elec. Dev. Lett.* **25**, 408–410 (2004).
19. Ritenour, A. *et al.* Epitaxial strained germanium p-MOSFETs with HfO₂ gate dielectric and TaN gate electrode. *IEDM Tech. Dig.*, 433–436 (2003).
20. Åberg, I., Ní Chléirigh, C., Olubuyide, O. O., Duan, X. & Hoyt, J. L. High electron and hole mobility enhancements in thin-body strained Si/strained SiGe/strained Si heterostructures on insulator. *IEDM Tech. Dig.*, 173–176 (2004).
21. Wallace, R. M. & Wilk, G. High- κ gate dielectric materials. *MRS Bull.* **27**, 192–197 (2002).
22. Wang, J., Polizzi, E., Ghosh, A., Datta, S. & Lundstrom, M. A theoretical investigation of surface roughness scattering in silicon nanowire transistors. *Appl. Phys. Lett.* **87**, 043101 (2005).
23. Wang, J. & Lundstrom, M. Does source-to-drain tunneling limit the ultimate scaling of MOSFETs? *IEDM Tech. Dig.*, 707–710 (2002).
24. Yeo, Y. C., King, T. J. & Hu, C. M. Metal-dielectric band alignment and its implications for metal gate complementary metal-oxide-semiconductor technology. *J. Appl. Phys.* **92**, 7266–7271 (2002).
25. Martel, R. *et al.* Ambipolar electrical transport in semiconducting single-wall carbon nanotubes. *Phys. Rev. Lett.* **87**, 256805 (2001).
26. Javey, A. *et al.* Carbon nanotube field-effect transistors with integrated ohmic contacts and high- κ gate dielectrics. *Nano Lett.* **4**, 447–450 (2004).
27. Lin, Y. M., Appenzeller, J. & Avouris, P. Ambipolar-to-unipolar conversion of carbon nanotube transistors by gate structure engineering. *Nano Lett.* **4**, 947–950 (2004).
28. Guo, J., Wang, J., Polizzi, E., Datta, S. & Lundstrom, M. Electrostatics of nanowire transistors. *IEEE Trans. Nanotechnol.* **2**, 329–334 (2003).
29. Patolsky, F. & Lieber, C. M. Nanowire nanosensors. *Mater. Today* **8**, 20–28 (2005).
30. Guo, J., Lundstrom, M. & Datta, S. Performance projections for ballistic carbon nanotube field-effect transistors. *Appl. Phys. Lett.* **80**, 3192–3194 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Y. Wen for help with cross-sectional TEM study and M. Radosavljevic for discussions. C.M.L. acknowledges support of this work by the Defense Advanced Projects Research Agency and Intel.

Author Contributions J.X., W.L., Y.H., Y.W. and H.Y. performed the experiments. J.X. and W.L. performed data analyses. J.X., W.L. and C.M.L. designed the experiments, discussed the interpretation of results and co-wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.M.L. (cml@cmliris.harvard.edu).

LETTERS

Long-term eruptive activity at a submarine arc volcano

Robert W. Embley¹, William W. Chadwick, Jr^{1,2}, Edward T. Baker³, David A. Butterfield^{3,4}, Joseph A. Resing^{3,4}, Cornel E.J. de Ronde⁵, Verena Tunnicliffe⁶, John E. Lupton¹, S. Kim Juniper⁷, Kenneth H. Rubin⁸, Robert J. Stern⁹, Geoffrey T. Lebon^{3,4}, Ko-ichi Nakamura¹⁰, Susan G. Merle^{1,2}, James R. Hein¹¹, Douglas A. Wiens¹² & Yoshihiko Tamura¹³

Three-quarters of the Earth's volcanic activity is submarine, located mostly along the mid-ocean ridges, with the remainder along intraoceanic arcs and hotspots at depths varying from greater than 4,000 m to near the sea surface. Most observations and sampling of submarine eruptions have been indirect, made from surface vessels or made after the fact^{1–6}. We describe here direct observations and sampling of an eruption at a submarine arc volcano named NW Rota-1, located 60 km northwest of the island of Rota (Commonwealth of the Northern Mariana Islands). We observed a pulsating plume permeated with droplets of molten sulphur disgorging volcanic ash and lapilli from a 15-m diameter pit in March 2004 and again in October 2005 near the summit of the volcano at a water depth of 555 m (depth in 2004). A turbid layer found on the flanks of the volcano (in 2004) at depths from 700 m to more than 1,400 m was probably formed by mass-wasting events related to the eruption. Long-term eruptive activity has produced an unusual chemical environment and a very unstable benthic habitat exploited by only a few mobile decapod species. Such conditions are perhaps distinctive of active arc and hotspot volcanoes.

Direct observations of submarine volcanic eruptions have only been made when their vents or products reach the ocean surface, as happened during eruptions of Myojinsho⁷, Surtsey⁸, Macdonald seamount² and Kavachi⁹. Such shallow eruptions are usually too violent to study with existing technology, although some samples of fluids, microbes and volcanic products were obtained from the eruption of Macdonald seamount in the late 1980s^{1,2,10}. Conversely, the more effusive deep-ocean basalt eruptions along the mid-ocean ridge have eluded direct observation. However, seafloor observations of fresh lavas, and sampling of fluids and microbes have been made within days to weeks following at least four deep submarine eruptions¹¹. These eruptions have been relatively small ($\leq 55 \times 10^6 \text{ m}^3$)¹² and short-lived (hours to days)^{13,14}. Phenomena associated with these mid-ocean ridge eruptions include: (1) large increases in volatiles produced by magma degassing and subsurface phase separation of the hydrothermal fluids^{6,15}, (2) blooms in microbial productivity driven by the increased volatile flux^{4,16,17}, and (3) in some cases, extinction and subsequent rapid renewal of the macrofaunal vent communities^{18,19}.

Submarine volcanism associated with intraoceanic island arcs

is quite different in character from mid-ocean ridge volcanism. Volatiles released during the subduction of an oceanic plate causes partial melting of the overlying mantle wedge, yielding lavas that are, on average, more siliceous and gas-rich than mid-ocean ridge basalt²⁰. These volcanoes are fixed with respect to their magma source for longer periods than their mid-ocean ridge counterparts, allowing their growth into shallow water or emergence as islands. The shallower depth range and the more explosive nature of submarine arc volcanism results in the generation of a much larger proportion of volcanoclastic material than at the mid-ocean ridge. These volatile-rich habitats on shallow isolated peaks may support fauna capable of surviving on intermittent hydrothermal production²¹.

A survey of more than 50 submarine volcanoes along the Mariana arc in February and March 2003 identified 12 sites with hydrothermal plumes²². The plume overlying the summit of NW Rota-1 (Fig. 1a) was distinguished by high rise height, elevated $\delta^3\text{He}$ values and low pH (ref. 22). Dives with the remotely operated vehicles (ROV) ROPOS and Hyper-Dolphin investigated NW Rota-1 in March/April 2004 and October 2005. NW Rota-1 is conical and about 16 km in diameter at its base at 2,700 m. The summit at 517 m lies along a ridge between a pair of southwest–northeast trending, inward-facing fault scarps that cut across the volcano. Rock samples collected during the dives at NW Rota-1 are vesicular, moderately fractionated, medium-K basalt and basalt andesites ($51.8 \pm 0.5\% \text{ SiO}_2$, $0.62 \pm 0.02\% \text{ K}_2\text{O}$, $6.4 \pm 1.1\% \text{ MgO}$)²³.

The south flank of the volcano between 700 and 884 m (Figs 1 and 2) is dominated by volcanoclastic debris and talus. Large-scale mass wasting is ubiquitous, with narrow spurs of lava forming headwalls of massive rock slides. The northwest–southeast trending summit ridge is composed primarily of volcanoclastic sand with some rock outcrop. Its western end has a sharp crest with a steep, unstable southern slope and a gentler northern slope (Fig. 3a). Deposits of black ash and lapilli (defined as volcanic particles $< 2 \text{ mm}$ and $2\text{--}64 \text{ mm}$ in diameter, respectively) intermixed with millimetre-sized sulphur globules were observed on flat surfaces or within small depressions on the outcrops along the summit ridge (Fig. 3b).

ROPOS surveys discovered an active crater $\sim 15 \text{ m}$ in diameter and $> 20 \text{ m}$ deep at a depth of 540 m on the south side of the summit. 'Brimstone Pit' (Fig. 2) was discharging a pulsating, opaque, yellowish smoky plume with characteristics unlike any known hydrothermal

¹NOAA/PMEL, 2115 SE O.S.U. Drive, Newport, Oregon 97365-5258, USA. ²CIMRS, Oregon State University, Oregon 97365-5258, USA. ³NOAA/PMEL, 7600 Sand Pt Way, NE, Seattle, Washington 98115-6349, USA. ⁴JISAO, University of Washington, Washington 98115-6349, USA. ⁵Institute of Geological and Nuclear Sciences, 30 Gracefield Road, PO Box 31-312, Lower Hutt, New Zealand. ⁶Department of Biology/School of Earth and Ocean Sciences, University of Victoria, PO Box 3020, Victoria, British Columbia V8W 3N5, Canada. ⁷GEOTOP Research Centre; Université du Québec à Montréal, PO Box 8888 Montréal, Québec H3C 3P8, Canada. ⁸Department of Geology and Geophysics, University of Hawaii, 1680 East-West Road, Honolulu, Hawaii 96822, USA. ⁹Geosciences Department, University of Texas at Dallas, 2601 N. Floyd Road, Richardson, Texas 75083-0688, USA. ¹⁰National Institute of Advanced Industrial Science and Technology, AIST Tsukuba Central 7, 1-1, Higashi 1-Chome Tsukuba, Ibaraki 305-8567, Japan. ¹¹US Geological Survey, 345 Middlefield Road, MS 999, Menlo Park, California 94025-3591, USA. ¹²Department of Earth and Planetary Sciences, Washington University, 1 Brookings Drive, St Louis, Missouri 63130-4899, USA. ¹³(IFREE) Japan Agency for Marine-Earth Science and Technology, 237-0061 2-15 Natsushima-chou, Yokosuka-shi, Kanagawa-ken, Japan.

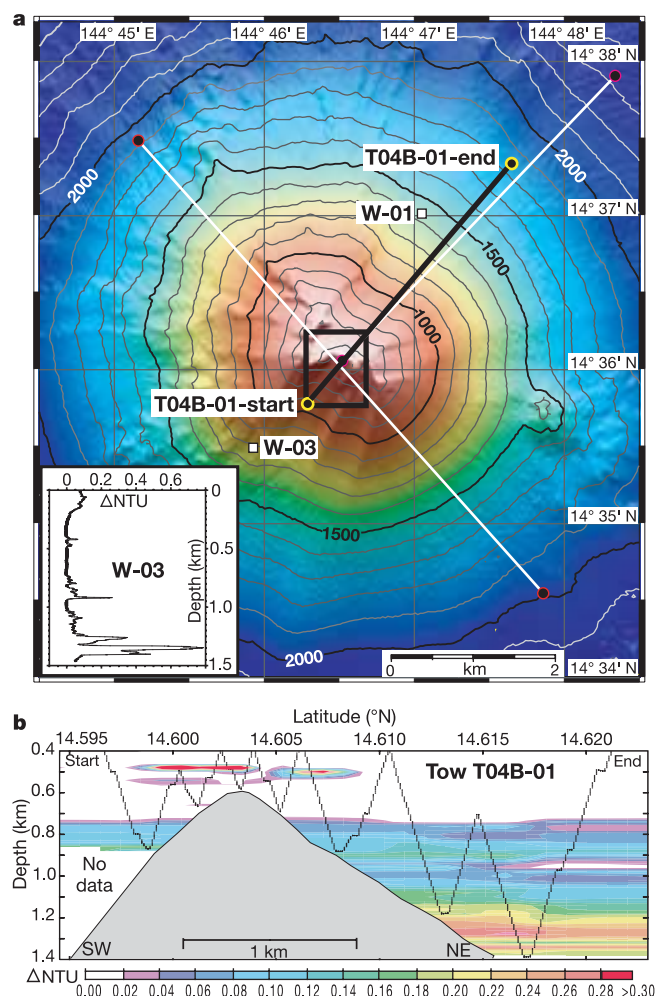


Figure 1 | Bathymetry of NW Rota-1 volcano and hydrothermal plumes.

a, EM300 multibeam bathymetry of upper portion of NW Rota-1 volcano, southern Mariana arc (grid-cell size is 30 m; contour interval is 100 m). The base of the volcano is approximately 2,770 m, beyond the map boundary. The area shown in Fig. 2 is indicated by the black box. CTD tows that encountered the deep turbidity plume are shown by straight lines and RV Wecoma vertical casts (W-01, W-03), taken six weeks after tows, are indicated by square symbols. The inset is the turbidity profile for W-03. Units are in nondimensional nephelometric turbidity units above the local ambient water (ΔNTU). **b**, Southwest to northeast turbidity profile (TO4B-01) over NW Rota-1. The scale bar at the bottom of the image is colour-coded in ΔNTU .

plumes (Fig. 3c). First, it exhibited extreme fluctuations in volume and intensity. Over a period of minutes explosive surges enveloped ROPOS before receding. Second, the plume had a distinct yellow hue that resembled sulphurous plumes emitted during volcanic bursts from subaerial volcanoes (for example, the White Island volcano offshore of New Zealand). Upon recovery, ROPOS was coated with thousands of solidified globules of sulphur (Fig. 3d). Their teardrop shape indicated a semi-molten state upon contact with the vehicle (the melting point of amorphous sulphur is 119 °C). Third, during the most intense discharges, lapilli and ash were ejected up and away from the pit. Radiometric dating using the ^{210}Po – ^{210}Pb method²⁴ of one of these fresh lava fragments that landed on ROPOS yielded a maximum eruption age of 3 March 2004, just 27 ± 12 days before sample collection on 30 Mar (see Methods). During a return visit to the site in October 2005 volcanic ‘bombs’ (at least 15 cm in diameter) were observed explosively degassing near the rim of a pit, showering the vehicle with volcanic debris (Fig. 3e). Some of these dark particles were observed in the plume about 25 m above the summit of the

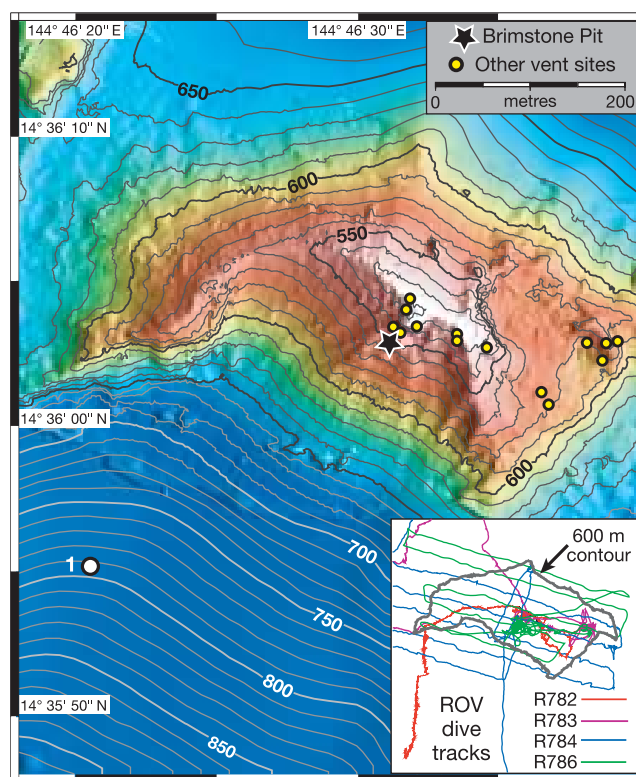
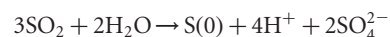


Figure 2 | Bathymetry and location map of the summit ridge of NW Rota-1 volcano with locations of features discussed in text. Contours are 10 m. High-resolution Imagenex sonar data over the ridge crest (2-m grid cell spacing) are overlaid on lower-resolution EM300 multibeam data. The white circle (point 1) is where extremely turbid layer was observed at the beginning of Dive R782. The lower-right inset shows the locations of dive tracks for R782–R784 and R786. Imagenex survey lines are the straighter northwest–southeast trending tracks.

volcano and 50 m above the rim of Brimstone Pit. This chronic fallout of volcanoclastic ejecta and sulphur globules is the origin of the deposits mantling the summit outcrops and the unstable slopes of the upper flank of the volcano.

Water sampled by ROPOS from ~5 m below the crater rim within Brimstone Pit during a lull in the activity in 2004 was thick with particulate sulphur, had dissolved sulphate concentration at least 15% above ambient background sea water, had low concentrations of H_2S ($<100 \mu\text{mol litre}^{-1}$) and was extremely acidic with a pH (measured at 25 °C) of around 2.0. The overall chemical composition of the sampled water and the sulphur isotope ratios²⁵ of hydrogen sulphide, elemental sulphur, and sulphate from Brimstone Pit are consistent with magmatic degassing of SO_2 and its disproportionation into sulphuric acid and elemental sulphur:



The water sampled from Brimstone Pit had a temperature as high as 30 °C in a zone of vigorous mixing. However, because this water was in contact with molten sulphur at a minimum of 119 °C, hotter water probably existed at the bottom of the pit.

A deeper dive (R782) encountered a highly turbid layer beginning at 700 m and continuing to the sea floor at a depth of 890 m (point 1, Fig. 2). A nephelometer on the CTD (a conductivity, temperature and depth measuring instrument) towed over the volcano during the 2004 expedition detected a zone of turbid layers surrounding the volcano from a depth of 700 m to over 1400 m (Fig. 1b). The scanning electron microscope imagery of particulate samples from the plume show that its suspended fraction consisted primarily of volcanic glass fragments²⁶. The deep turbid layer was uniquely

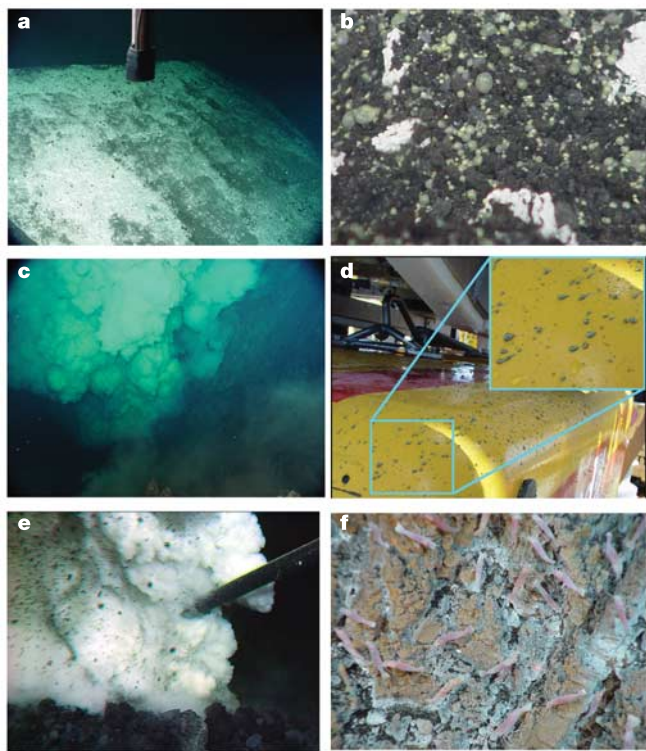


Figure 3 | Images of the sea floor at NW-Rota-1. **a**, The summit ridge showing the unstable south flank of volcaniclastic sand. The image is ~5 m across. **b**, Close-up of sea floor near Brimstone Pit composed of sulphur globules (pale spheres ~2–3 mm across), black lapilli and ash and unknown white material. **c**, A cloud of sulphur-rich fluid erupting from Brimstone Pit in March 2004. The image is ~5 m across. Video clips of the 2004/2005 activity in Brimstone Pit can be viewed at: <http://www.oceanexplorer.noaa.gov/explorations/04fire/logs/brimstone/brimstone.html>. **d**, Sulphur droplets coat ROPOS after dive R784 to Brimstone Pit. **e**, Showers of particles from explosive degassing of volcanic bombs at pit, October 2005. Frame-grab of high-definition video on Hyper-Dolphin remotely operated vehicle. **f**, Shrimp (~2–3 cm long) swarming on an outcrop near Brimstone Pit.

associated with NW Rota-1; a CTD cast made at 14°32.13' N, 144°51.78' (~12 km southeast of NW Rota-1) encountered no significant turbidity anomaly. Two CTD casts (W-01, W-03) made by the research vessel RV *Wecoma* six weeks after the RV *T. G. Thompson* measurements found a somewhat diminished deep turbidity layer on the northeast and southwest flanks of the volcano (Fig. 1a, inset). The deep turbid layer was not present on full water depth CTD profiles taken in 2003 (ref. 22), nor was the high turbidity visually apparent on the video imagery in 2005 even though the summit eruptive activity was more intense. The origin of the deep turbid layer is uncertain, although periodic collapse of loose volcaniclastic material built up from chronic eruptive activity (for example, an apparent shoaling of the eruption site by 20 m occurred between April 2004 and October 2005) is a likely mechanism.

Vent-associated animals coexisted with microbial mats at diffusely venting sites above Brimstone Pit across to the eastern ridge (Fig. 2). The abundance of crabs was low (*Austinograea cf. yunohana*) but two species of alvinocaridid shrimp (*Opaepele loihi* and an undescribed species (under description by R. Webb)) reached localized densities of over 200 individuals per m² (Fig. 3f). Exposed sessile animals that occur on the inactive peak a few hundred metres to the northwest and elsewhere at Mariana and Izu-Bonin arc vents (observations made during the 2004 RV *T. G. Thompson* Submarine 'Ring of Fire' expedition by V.T. and S.K.J. in 2004) were absent; a few limpets occurred below overhangs. This assemblage was probably persistent

despite the chronic eruptive disruptions: complete size ranges of juveniles and adults of both shrimp species were present, indicating continuous recruitment, and the same assemblage was observed twice, more than a year apart. Sustained eruptive activity with chronic fallout of volcanic ejecta, congealing sulphur and unstable surfaces probably exclude extensive colonization of the hydrothermal outlets by sessile fauna.

The unusual physical and chemical character of the Brimstone Pit and environs, the unstable volcanoclastic slopes, the juvenile ejecta, the mobile biologic community and the turbid plume around the lower flank of the volcano observed in 2004 indicate that we were observing chronic eruptive activity at NW Rota-1. Submarine eruption plumes have been classified by the ratio of mass transport of water to the mass flux of magma^{27,28}. The Brimstone Pit plume at NW Rota-1, at least during the periods of our observations, was characterized by relatively high concentrations of magmatic gas, low magma throughput and high seawater entrainment. This results in the production and fallout of sulphur precipitates and volcanic material near the vent and occasional volcanoclastic density flows at an arc volcano. This study provides the first observations of a deep submarine eruption of an arc volcano, but this case is just one example of a wide range of phenomena that needs to be observed and sampled to understand the behaviour of submarine arc volcanoes better. The persistence of eruptive activity for at least 18 months at NW Rota-1 also implies that it will be possible to learn much about the evolution of submarine arc volcanoes and the nature of their chemical, geological and biologic systems by *in situ* monitoring.

METHODS

²¹⁰Po–²¹⁰Pb dating. Very fresh glass was handpicked under a binocular microscope for ²¹⁰Po–²¹⁰Pb dating of the lava fragment collected at NW Rota-1 in 2004. The sample was dissolved and ²¹⁰Po (half life = 138.4 day) was repeatedly determined by α -spectrometry in aliquots over one year (ref. 24). The eruption age was determined by best-fit regression to a radioactive ingrowth curve²⁴. Polonium volatilization from erupting magmas creates an initial ²¹⁰Po deficit relative to grand parental ²¹⁰Pb in freshly erupted magmas, which is subsequently erased with time via radioactive ingrowth towards secular equilibrium. Incomplete Po degassing would cause reported ages to be overestimates, and so the reported age is nominally a maximum. A sample from shallow submarine Macdonald seamount¹ was 100% degassed on the date of sample collection (presumed eruption), yet degassing at greater depth (~3 km) is presently constrained to just 75–100% (ref. 24). The lapilli fragment dated here contained 13% of its pre-eruptive ²¹⁰Po on the date of sample collection, indicating that at least 87% of the Po was lost upon eruption. The shallow depth and copious production of S aerosols implies that the maximum age (100% degassing) is a realistic estimate. Error in maximum age is a worst-case scenario determined by regression through the data at the limits of their analytical error ranges.

Received 3 October 2005; accepted 22 March 2006.

1. Rubin, K. H. & Macdougall, J. D. Submarine magma degassing and explosive magmatism at Macdonald (Tamarii) seamount. *Nature* **341**, 50–52 (1989).
2. Cheminee, J.-L. *et al.* Gas-rich submarine exhalations during the 1989 eruption of Macdonald Seamount. *Earth Planet. Sci. Lett.* **107**, 318–327 (1991).
3. Baker, E. T., Massoth, G. J. & Feely, R. A. Cataclysmic venting on the Juan de Fuca Ridge. *Nature* **329**, 149–151 (1987).
4. Haymon, R. M. *et al.* Volcanic eruption of the mid-ocean ridge along the East Pacific Rise crest at 9°45'–52' N: Direct submersible observations of seafloor phenomena associated with an eruption event in April, 1991. *Earth Planet. Sci. Lett.* **119**, 85–101 (1993).
5. Embley, R. W., Chadwick, W. W., Jonasson, I. R., Butterfield, D. A. & Baker, E. T. Initial results of the rapid response to the 1993 CoAxial event: Relationships between hydrothermal and volcanic processes. *Geophys. Res. Lett.* **22**, 143–146 (1995).
6. Butterfield, D. *et al.* Seafloor eruptions and evolution of hydrothermal fluid chemistry. *Phil. Trans. R. Soc. Lond. A* **355**, 369–386 (1997).
7. Fiske, R. S., Cashman, K. V., Shibata, A. & Watanabe, K. Tephra dispersal from Myojinsho, Japan, during its shallow submarine eruption of 1952–1953. *Bull. Volcanol.* **59**, 262–275 (1998).
8. Kokelaar, B. P. & Durant, G. P. The submarine eruption and erosion of Surtla (Surtsey), Iceland. *J. Volcanol. Geotherm. Res.* **19**, 239–246 (1983).

9. Baker, E. T., Massoth, G. J., de Ronde, C. E. J., Lupton, J. E. & McInnes, B. I. A. Observations and sampling of an ongoing subsurface eruption of Kavachi volcano, Solomon Islands, May 2000. *Geology* **30**, 975–978 (2002).
10. Huber, R., Stoffer, P., Cheminee, J.-L., Richnow, H. H. & Stetter, K. O. Hyperthermophilic archaeobacteria within the crater and open-sea plume of erupting Macdonald Seamount. *Nature* **345**, 179–182 (1990).
11. Cowen, J. P., Baker, E. T. & Embley, R. W. in *The Seafloor Biosphere at Mid-Ocean Ridges* (eds Wilcock, W. S. D., Cary, C., DeLong, E., Kelley, D. S. & Baross, J. A.) 227–243 (AGU Geophysical Monograph 144, American Geophysical Union, Washington DC, 2004).
12. Chadwick, W. W., Embley, R. W. & Shank, T. M. The 1996 Gorda Ridge eruption: geologic mapping, sidescan sonar, and SeaBeam comparison results. *Deep-Sea Res.* **45**, 2547–2569 (1998).
13. Gregg, T. K. P. & Fink, J. H. Quantification of submarine lava-flow morphology through analog experiments. *Geology* **23**, 73–76 (1995).
14. Fox, C. G., Chadwick, W. W. Jr & Embley, R. W. Direct observation of a submarine volcanic eruption from a sea-floor instrument caught in a lava flow. *Nature* **412**, 727–729 (2001).
15. Von Damm, K. L. *et al.* Evolution of East Pacific Rise hydrothermal vent fluids following a volcanic eruption. *Nature* **375**, 47–50 (1995).
16. Delaney, J. R. *et al.* The quantum event of crustal accretion: Impacts of diking at Mid-Ocean Ridges. *Science* **281**, 222–230 (1998).
17. Embley, R. W. & Lupton, J. E. in *The Seafloor Biosphere at Mid-Ocean Ridges* (eds Wilcock, W. S. D., Cary, C., DeLong, E., Kelley, D. S. & Baross, J. A.) 75–97 (AGU Geophysical Monograph 144, American Geophysical Union, Washington DC, 2004).
18. Shank, T. M. *et al.* Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50' N, East Pacific Rise). *Deep-Sea Res. II* **45**, 465–515 (1998).
19. Tunnicliffe, V. *et al.* Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca Ridge. *Deep-Sea Res.* **44**, 1627–1644 (1997).
20. Stern, R. J. Subduction zones. *Rev. Geophys.* **40**, doi:10.1029/2001RG000108 (2002).
21. Karl, D. M., McMurtry, G. M., Malahoff, A. & Garcia, M. O. Loihi Seamount, Hawaii: A mid-plate volcano with a distinctive hydrothermal system. *Nature* **335**, 532–534 (1988).
22. Embley, R. W. *et al.* Explorations of Mariana arc volcanoes reveal new hydrothermal systems. *Eos* **85**, 37–40 (2004).
23. Stern, R. J., Basu, N. K., Kohut, E., Hein, J. & Embley, R. W. Petrology and geochemistry of igneous rocks collected in association with ROV investigations of three hydrothermal sites in the Mariana Arc: NW Rota-1, E. Diamante, and NW Eifuku. *Eos (Fall Meet. Suppl.)* **85**, abstr. V43F–07 (2004).
24. Rubin, K. H., Macdougall, J. D. & Perfit, M. R. ²¹⁰Po–²¹⁰Pb dating of recent volcanic eruptions on the sea floor. *Nature* **368**, 841–844 (1994).
25. Miyabe, S. *et al.* Sulfur isotope geochemistry of Mariana arc hydrothermal systems. *Eos (Fall Meet. Suppl.)* **85**, abstr. V41B–1387 (2004).
26. Lebon, G. T., Resing, J. E., Baker, E. T., Embley, R. W. & Lupton, J. E. Chemical characteristics of plumes above and around NW-Rota submarine volcano: Before and during the March 2004 eruption. *Trans. AGU* **85**, abstr. V41B–1386 (2004).
27. Koyaguchi, T. & Woods, A. W. On the formation of eruption columns following explosive mixing of magma and surface water. *J. Geophys. Res.* **101**, 5561–5574 (1996).
28. White, J. D. L., Smellie, J. L. & Clague, D. A. in *Explosive Subaqueous Volcanism* (eds White, J. D. L., Smellie, J. L. & Clague, D. A.) 1–23 (American Geophysical Union, Washington DC, 2003).

Acknowledgements The 2004 Submarine Ring of Fire Program was supported by the NOAA Office of Ocean Exploration, the Canadian Natural Sciences and Engineering Research Council, and the NOAA VENTS Program. The MARGINS Program of the National Science Foundation provided support for the May 2004 CTD operations from the RV *Wecoma*. We are grateful for the support by the RV *Thomas G. Thompson*, operated by the University of Washington, the Canadian Scientific Submersible Scientific Support Facility who operated the ROPOS ROV on the 2004 expedition, and the Japan Agency for Marine, Earth Science and Technology (JAMSTEC) who operate the RV *Natsushima* and the Hyper-Dolphin ROV and, funded the Japanese science team in 2005 (cruise NT-05-17). S. Ristau, R. Greene, S. Walker, L. Evans and K. Roe provided support at sea. We also thank our colleagues T. Urabe and J. Ishibashi for their collaboration, which helped make the 2004 expedition to the Mariana Arc possible. This publication is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under an NOAA Cooperative Agreement and by the Pacific Marine Environmental Laboratory.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.E. (robert.w.embley@noaa.gov).

LETTERS

Resource competition and social conflict in experimental populations of yeast

R. Craig MacLean¹ & Ivana Gudelj²

Understanding the conditions that promote the maintenance of cooperation is a classic problem in evolutionary biology^{1–5}. The essence of this dilemma is captured by the ‘tragedy of the commons’⁶: how can a group of individuals that exploit resources in a cooperative manner resist invasion by ‘cheaters’ who selfishly use common resources to maximize their individual reproduction at the expense of the group^{7,8}? Here, we investigate this conflict through experimental competitions between isogenic cheater and cooperator strains of yeast with alternative pathways of glucose metabolism⁹, and by using mathematical models of microbial biochemistry¹⁰. We show that both coexistence and competitive exclusion are possible outcomes of this conflict, depending on the spatial and temporal structure of the environment. Both of these outcomes are driven by trade-offs between the rate and efficiency of conversion of resources into offspring that are mediated by metabolic intermediates.

Heterotrophic microorganisms depend on metabolic pathways to convert resources, such as sugars, into energy in the form of ATP, which is required for growth and reproduction. Trade-offs between the rate and yield of substrate degradation set the stage for evolutionary conflict between individuals with alternative metabolic pathways that convert resources into energy either rapidly or efficiently^{7,8}. Efficient use of common resources conforms to the classical definition of a cooperative trait: it is beneficial to the group, because more biomass is produced per unit resource, but costly to individuals, because each cell produces energy and divides at a slow rate.

Under what conditions can a group of cooperating microbial cells resist invasion by cheaters who wastefully deplete common resources to increase their growth rate? Theoretical studies have addressed this question by modelling competition between cooperative respirers that take up glucose slowly and fully respire all of the glucose they ingest, resulting in a high yield of ATP production, and selfish respiration-fermenters that take up glucose rapidly and ferment any excess glucose that cannot be respired, resulting in a high rate of ATP production^{7,8,11}. In these models the outcome of competition depends on the scale of resource competition: cheaters exclude cooperators when competition is for a global pool of resources in a homogeneous environment; cooperators resist invasion by cheaters if competition occurs for local resource patches in a spatially structured environment^{7,8,11}.

To investigate the outcome of global competition, we competed isogenic respirer (cooperator) and respiration-fermenter (cheater) yeast strains^{9,12} in glucose-limited chemostats. The cheater has a higher fitness than the cooperator in the chemostat (mean $w = 1.08$, s.e.m. = 0.022, $t_{(1),11} = 3.43$, $P = 0.0028$) and competition results in exclusion of the cooperator. To investigate the effect of seasonality on global competition, we carried out a second set of experiments in glucose-limited batch cultures. Fitness in batch cultures is negative

frequency-dependent ($F_{1,46} = 58$, $P < 0.0001$, $r^2 = 0.59$; Fig. 1a) and both strains are mutually invisable, implying that the outcome of competition is stable coexistence. Competitive fitness of the cooperator in batch cultures is also positive density-dependent ($F_{1,18} = 12.3$, $P = 0.002$; Fig. 1b), implying that the cooperator achieves a higher equilibrium frequency when population density at the start of a season is high.

To investigate mechanisms of coexistence, we developed a seasonal model of competition for a sugar. For simplicity, we model the catabolism of sugar and its intermediates as a two-reaction process¹⁰ corresponding to glycolysis and the tricarboxylic acid (TCA) cycle. In

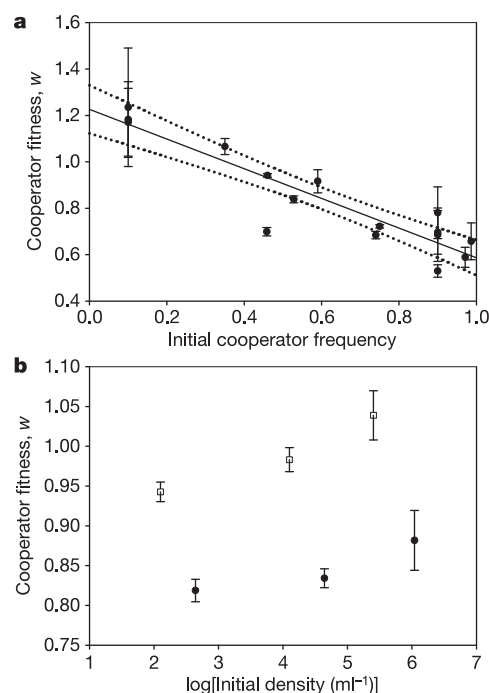


Figure 1 | Seasonal competition between cooperators and cheaters.

a, Plotted points show cooperator fitness ($\pm$ s.e.m.; $n = 3$ or 4) as a function of initial frequency. The solid line shows the regression of fitness on frequency and dotted lines show the 95% confidence interval about this regression. Initial glucose concentration varied between competitions, but the slope ($F_{1,12} = 0.79$, $P = 0.39$) and intercept ($F_{1,12} = 0.4$, $P = 0.54$) of the regression are independent of initial glucose concentration. Initial population density was constant in all competitions. **b**, Plotted points show the competitive fitness of the cooperator strain ($\pm$ s.e.m.; $n = 4$) at an initial frequency of 0.9 (closed symbols) or 0.4 (open symbols) across a gradient of initial population density. Initial glucose concentration was the same in all competition.

¹NERC Center for Population Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK. ²Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, UK.

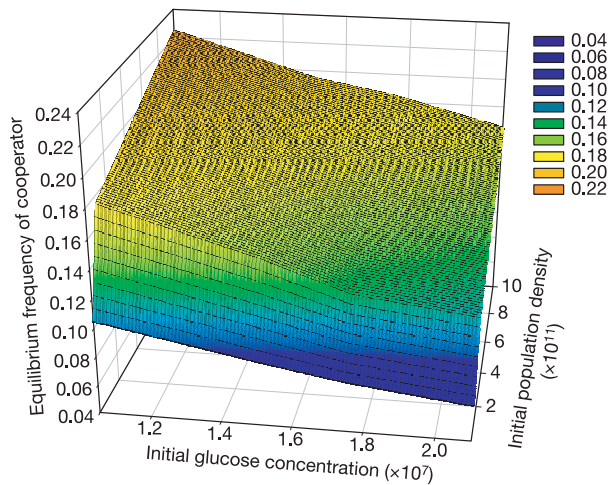


Figure 2 | Modelling competition between cheaters and cooperators. This plot shows the outcome of competition across different initial densities and glucose concentrations. The frequency of the cooperator strain at steady state is plotted against different initial densities and glucose concentrations (in nanomoles) at the beginning of each season.

the first reaction, sugar is taken from the environment and partially oxidized to form an intracellular metabolic intermediate and n_1 molecules of ATP. In the second reaction, the intracellular intermediate is either completely oxidized to form CO_2 and n_2 molecules of ATP, or it passively diffuses out of the cell as an extracellular intermediate. We assume that both catabolic reactions show saturating enzyme kinetics, that the rate of cell growth is proportional to the rate of ATP production¹³ according to a proportionality constant G , and that the intracellular intermediate imposes a cost in terms of lost ATP given by a linear function g as a result of deleterious effects of metabolites such as end-product inhibition and direct toxicity^{10,14} (See Supplementary Fig. 1 for the pathway schematic). We parameterized this model using data on the biochemistry of yeast, assuming that the only innate difference between strains is that cheaters have a higher maximum rate of glycolysis (Supplementary Table 1). The dynamics of this model within one season are described as follows, where S is the concentration of sugar in moles, X_{in} and X_{ex} are the concentrations of the intracellular and extracellular intermediates in moles respectively, v_1 and v_2 are the rate of glycolysis and the TCA cycle respectively, c is a transport constant, N denotes cell density, and the subscripts r and f denote cooperators and cheaters (units of measurement are presented in Supplementary Table 1):

$$\begin{aligned} dS/dt &= -v_{1r}(S)N_r - v_{1f}(S)N_f \\ dX_{ex}/dt &= c[X_{in,r} - X_{ex}]N_r + c[X_{in,f} - X_{ex}]N_f \\ dN_r/dt &= Gg(X_{in,r})[v_{1r}(S)n_1 + v_2(X_{in,r})n_2]N_r \\ dN_f/dt &= Gg(X_{in,f})[v_{1f}(S)n_1 + v_2(X_{in,f})n_2]N_f \\ dX_{in,r}/dt &= [-c[X_{in,r} - X_{ex}] + v_{1r}(S) - v_2(X_{in,r})]N_r \\ dX_{in,f}/dt &= [-c[X_{in,f} - X_{ex}] + v_{1f}(S) - v_2(X_{in,f})]N_f \end{aligned} \quad (1)$$

This model predicts density-dependent coexistence of both strategies in seasonal environments at frequencies similar to those observed in our competition experiments, implying that cheating carries costs and benefits (Fig. 2). The benefit of fermentation is that it gives cheaters a higher rate of ATP production for any given value of S , resulting in a high maximum growth rate⁹. The cost of this strategy is

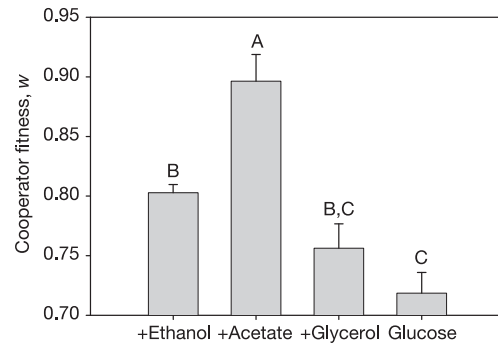


Figure 3 | Toxic metabolic intermediates create a cost of cheating.

Cooperator and cheater strains were competed against each other on agar plates containing 8 g l^{-1} glucose or 8 g l^{-1} glucose supplemented with glycerol, acetate, or ethanol at a concentration of 4 g l^{-1} each. The initial population density and frequency of the cooperator were the same in all competitions. Bars show the mean competitive fitness of the cooperator strain ($\pm \text{s.e.m}$; $n = 3$). Mean fitness differs among treatments that are not connected by the same letter (A, B, C), as judged by a one-way analysis of variance, ANOVA ($F_{3,8} = 18.6$, $P = 0.0006$), followed by a Tukey test ($P < 0.05$).

twofold. Cheaters pay a cost in terms of reduced reproductive yield (number of offspring per mole of ATP), because they are prone to accumulate high concentrations of toxic intermediates (see Supplementary Note 1 for the mathematical explanation). This cost is high when extracellular intermediates accumulate to high concentrations, because X_{ex} inhibits the diffusion of excess X_{in} out of cheater cells. Consistent with this expectation, the addition of toxic metabolites (ethanol and acetate) to batch culture competitions increases cooperator fitness, while the addition of a non-toxic metabolite (glycerol) has no significant effect on fitness (Fig. 3). The evolutionary consequence of reduced reproductive yield is that the relative growth rate (that is, the fitness) of the cheater declines as intermediates accumulate. The second cost of cheating is that it lowers the biochemical yield of energy production (moles of ATP per mole of sugar). This cost can be illustrated as follows: cooperators always gain n_2 molecules of ATP from each molecule of X_{in} they produce, while cheaters gain less than n_2 molecules of ATP from each X_{in} they produce, because some of the excess intermediates produced by cheaters are subsequently oxidized by cooperators after S has been depleted.

Seasonality promotes coexistence in this system because it allows metabolic intermediates to accumulate, reducing the biochemical and reproductive yield of energy production in cells with a high biochemical rate (moles of ATP per unit of time) of energy production. Consistent with this argument, the fitness of the cheater is lowest under conditions that promote the accumulation of high concentrations of extracellular metabolites, namely high absolute cheater density (Fig. 1). This result is of general importance to our understanding of social conflict in microbial populations because trade-offs between the rate and yield of energy production are a general feature of metabolism that are not specific to this system^{7,8}. In this case, metabolic intermediates are the molecules that mediate this trade-off. Coexistence is not possible in the chemostat in this system because intermediates are constantly washed out of the chemostat, minimizing the cost of cheating (Supplementary Equation (1)), although both coexistence and competitive exclusion of cheaters are theoretically possible in chemostat models that assume much higher costs of cheating¹⁰.

Spatial structure promotes the evolution of cooperation in game theoretical models of social conflict because it allows synergistic interactions among localized populations of cooperators to result in large payoffs that are not accessible to cheaters^{5,8}. The benefit of

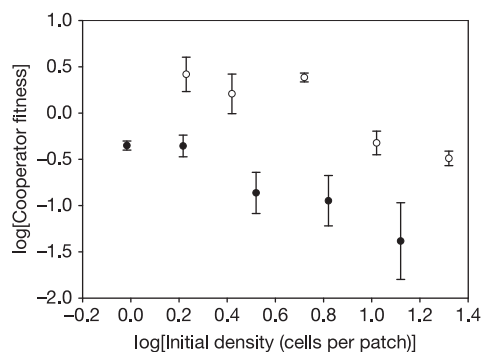


Figure 4 | Competition in a spatially heterogeneous environment. Plotted points show the mean fitness ($\pm$ s.e.m.; $n = 3$) of the cooperator strain at the level of the metapopulation. The initial frequency of the cooperator strain was either 0.91 (open symbols) or 0.45 (closed symbols). Because the number of generations of competition varied between patches in the metapopulations, fitness was not standardized to the number of generations of competition.

cooperation in this system is that a group of cooperators that exploit a local resource patch will produce more offspring than an equivalent population of cheaters because they have a higher reproductive yield of ATP production. In a spatially heterogeneous environment consisting of many local patches, cooperation will be favoured when (1) the frequency of the cheater is low, or (2) the number of colonists per patch is low, because many patches will be exclusively colonized by cooperators under these two scenarios.

To test this hypothesis, the cheater and cooperator strains were inoculated at random into a metapopulation made up of 96 discrete patches. After one season of growth, we pooled the populations from each patch in the metapopulation to assay the outcome of competition at the level of the metapopulation (Fig. 4). Within any given patch, competition is seasonal. In agreement with our hypothesis, the fitness of the cooperator at the level of the metapopulation is both positive frequency-dependent ($F_{1,7} = 55$, $P < 0.0001$; Fig. 4) and negative density-dependent ($F_{1,7} = 41$, $P < 0.0001$; Fig. 4), implying that a metapopulation of cooperators can resist invasion by a rare cheater provided that the number of cells that colonize each patch is below a critical threshold. Positive frequency-dependent selection for cooperation also implies that a rare cooperator cannot invade a population of cheaters.

Competition for common resources is one of the simplest social conflicts that can be imagined. Evolutionary game theory models based on the Prisoner's Dilemma predict that the outcome of this conflict is simple: cheating will prevail in homogeneous populations^{2,7,8,11} and spatial structure will promote the evolution of cooperation^{5,7,8}. This study reports two experimental findings that are of general relevance for our understanding of social conflict: (1) cooperation can persist in a well-mixed population in the absence of kin-recognition, policing or rational behaviour and (2) spatial structure can promote the maintenance of cooperation.

Social conflict is recognized as an important aspect of microbial population biology^{15–18} and metabolism^{17,19} even though microbes are not capable of rational thought. Both theoretical and experimental studies of these conflicts have investigated the fitness of fixed strategies, even though microbes are capable of rapid changes in gene expression in response to simple environmental cues such as substrate availability. The evolutionary optimal strategy in the conflict we describe here may be a plastic genotype that switches between selfish and efficient metabolism depending on the concentration of intermediates in the environment, instead of the 'wild-type' *Saccharomyces cerevisiae* strategy, which is to ferment any excess glucose that cannot be respired irrespective of ethanol concentration⁹. Future studies that investigate microbial social conflict at

the level of individual cells will hopefully shed light on this intriguing possibility.

METHODS

Yeast strains. We used yeast strains CEN.PK2-1C and TM6*, an isogenic respirer mutant that carries a single, synthetic hexose transporter gene⁹. Both strains are haploid and MATa.

Chemostat competitions. Strains were competed over two days in chemostats supplied with glucose-limited medium (glucose 0.8 g l⁻¹, yeast nitrogen base 1.7 g l⁻¹, ammonium sulphate 5 g l⁻¹, uracil 0.002%) that were incubated at 30 °C with continuous shaking (150 r.p.m.) and aeration. The dilution rate of the chemostats varied between 0.27 and 0.35 to ensure that that glucose flux was high enough to allow the cheater mutant to ferment²⁰.

Batch-culture competitions. Pre-inoculation cultures were mixed at various ratios to form common pools that were used to inoculate replicate competition cultures (glucose 0.25 g l⁻¹–20 g l⁻¹, yeast nitrogen base 1.7 g l⁻¹, ammonium sulphate 5 g l⁻¹), supplemented with uracil (0.002%) and agar (1.6%) where necessary. Competition cultures were incubated for two days at 30 °C; flasks were continuously shaken at a speed of 150 r.p.m.

Spatial competition. Common pools were inoculated into 96-well microplates containing 150 µl of competition medium (glucose 20 g l⁻¹, yeast nitrogen base 1.7 g l⁻¹, ammonium sulphate 5 g l⁻¹, uracil 0.002%). Microplates were incubated for 5 days at 30 °C without shaking.

Estimating fitness. The population density of the two strains at the start and end of competition was measured by plating appropriate serial dilutions on YPD agar plates. At low population density, the two strains form very large (CEN.PK2-1C) or small (TM6*) colonies. The accuracy of this technique was confirmed by replica-plating neutrally marked strains (ura⁻, kan^R) onto appropriate selective media. Competitive fitness was calculated as w per generation, where w is the exponent of the change in the natural logarithm of the ratio of the density of the two strains during the course of competition divided by the number of population doublings that occurred during competition, such that a value of $w = 1$ represents equal competitive ability.

Mathematical modelling. A population of N_c cooperator and N_f cheater cells was introduced into an environment containing a fixed initial concentration of resources and no metabolic intermediates. Competition occurred until all resources and intermediates were depleted from the environment. We assumed that only a proportion of the total population at the end of a season survived into the next season, in which the environment contained the same fixed resource concentration as in the previous season and no metabolic intermediates. The total cell density at the beginning of each season was kept constant and therefore the only component that changed at the beginning of each season was the cooperator/cheater ratio. The stable coexistence between cooperators and cheaters was said to be reached if the cooperator/cheater ratio remained constant between the seasons. Computer simulations were conducted using MATLAB while the model was parameterized with data on the biochemistry of the cooperator and cheater strains used in our competition experiments (Supplementary Information 1).

Received 15 December 2005; accepted 2 February 2006.

1. Axelrod, R. & Dion, D. The further evolution of cooperation. *Science* **242**, 1385–1390 (1988).
2. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
3. Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. The evolution of cooperation. *Q. Rev. Biol.* **79**, 135–160 (2004).
4. Nowak, M. & Sigmund, K. Chaos and the evolution of cooperation. *Proc. Natl Acad. Sci. USA* **90**, 5091–5094 (1993).
5. Nowak, M., Bonhoeffer, S. & May, R. Spatial games and the maintenance of cooperation. *Proc. Natl Acad. Sci. USA* **91**, 4877–4881 (1994).
6. Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
7. Kreft, J.-U. Biofilms promote altruism. *Microbiology* **150**, 2751–2760 (2004).
8. Pfeiffer, T., Schuster, S. & Bonhoeffer, S. Cooperation and competition in the evolution of ATP-producing pathways. *Science* **292**, 504–507 (2001).
9. Otterstedt, K. et al. Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. *EMBO Rep.* **5**, 532–537 (2004).
10. Pfeiffer, T. & Bonhoeffer, S. Evolution of cross-feeding in microbial populations. *Am. Nat.* **163**, E126–E135 (2004).
11. Frick, T. & Schuster, S. An example of the prisoner's dilemma in biochemistry. *Naturwissenschaften* **90**, 327–331 (2003).
12. Elbing, K. et al. Role of hexose transport in control of glycolytic flux in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **70**, 5323–5330 (2004).
13. Bauchop, T. & Elsdon, S. R. The growth of microorganisms in relation to their supply of energy. *J. Gen. Microbiol.* **23**, 457–469 (1960).
14. Ludovico, P., Sousa, M. J., Silva, M. T., Leao, C. & Corte-Real, M. *Saccharomyces*

- cerevisiae* commits to a programmed cell death process in response to acetic acid. *Microbiology* **147**, 2409–2415 (2001).
15. Travisano, M. & Velicer, G. J. Strategies of microbial cheater control. *Trends Microbiol.* **12**, 72–78 (2004).
 16. Crespi, B. J. The evolution of social behaviour in microorganisms. *Trends Ecol. Evol.* **16**, 178–183 (2001).
 17. Greig, D. & Travisano, M. The Prisoner's Dilemma and polymorphism in yeast *SUC* genes. *Proc. R. Soc. Lond. B* **271** (Suppl.), S25–S26 (2004).
 18. Velicer, G. J. Social strife in the microbial world. *Trends Microbiol.* **11**, 330–337 (2003).
 19. Pfeiffer, T. & Schuster, S. Game-theoretical approaches to studying the evolution of biochemical systems. *Trends Biochem. Sci.* **30**, 20–25 (2005).
 20. van Dijken, J. P. *et al.* An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* **26**, 706–714 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This project was funded by grants from NERC to the Center for Population Biology. I. Gudelj was supported by a NERC-EMS Fellowship. The authors would like to thank C. Godfray for comments and C. Brandon for technical assistance.

Author Contributions R.C.M. carried out experimental work and wrote the manuscript. I.G. carried out mathematical modelling.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.C.M. (c.maclea@imperial.ac.uk).

LETTERS

A simple rule for the evolution of cooperation on graphs and social networks

Hisashi Ohtsuki^{1,2}, Christoph Hauert², Erez Lieberman^{2,3} & Martin A. Nowak²

A fundamental aspect of all biological systems is cooperation. Cooperative interactions are required for many levels of biological organization ranging from single cells to groups of animals^{1–4}. Human society is based to a large extent on mechanisms that promote cooperation^{5–7}. It is well known that in unstructured populations, natural selection favours defectors over cooperators. There is much current interest, however, in studying evolutionary games in structured populations and on graphs^{8–17}. These efforts recognize the fact that who-meets-whom is not random, but determined by spatial relationships or social networks^{18–24}. Here we describe a surprisingly simple rule that is a good approximation for all graphs that we have analysed, including cycles, spatial lattices, random regular graphs, random graphs and scale-free networks^{25,26}: natural selection favours cooperation, if the benefit of the altruistic act, b , divided by the cost, c , exceeds the average number of neighbours, k , which means $b/c > k$. In this case, cooperation can evolve as a consequence of ‘social viscosity’ even in the absence of reputation effects or strategic complexity.

A cooperator is someone who pays a cost, c , for another individual to receive a benefit, b . A defector pays no cost and does not distribute any benefits. In evolutionary biology, cost and benefit are measured in terms of fitness. Reproduction can be genetic or cultural. In the latter case, the strategy of someone who does well is imitated by others. In an unstructured population, where all individuals are equally likely to interact with each other, defectors have a higher average payoff than unconditional cooperators. Therefore, natural selection increases the relative abundance of defectors and drives cooperators to extinction. These evolutionary dynamics hold for the deterministic setting of the replicator equation^{27,28} and for stochastic game dynamics of finite populations²⁹.

In our model, the players of an evolutionary game occupy the vertices of a graph. The edges denote links between individuals in terms of game dynamical interaction and biological reproduction. We assume that the graph is fixed for the duration of the evolutionary dynamics. Consider a population of N individuals consisting of cooperators and defectors. A cooperator helps all individuals to whom it is connected. If a cooperator is connected to k other individuals and i of those are cooperators, then its payoff is $bi - ck$. A defector does not provide any help, and therefore has no costs, but it can receive the benefit from neighbouring cooperators. If a defector is connected to j cooperators, then its payoff is bj .

The fitness of an individual is given by a constant term, denoting the baseline fitness, plus the payoff that arises from the game. Strong selection means that the payoff is large compared to the baseline fitness; weak selection means the payoff is small compared to the baseline fitness. The idea behind weak selection is that many different factors contribute to the overall fitness of an individual, and the game under consideration is just one of those factors.

At first, we will study the following update rule for evolutionary dynamics (Fig. 1): in each time step, a random individual is chosen to die, and the neighbors compete for the empty site proportional to their fitness. We call this mechanism ‘death–birth’ updating, because it involves a death event followed by a birth. Later we will investigate other update mechanisms.

Let us explore whether natural selection can favour cooperation on certain graphs. To do this, we need to calculate the probability that a single cooperator starting in a random position turns the whole population from defection to cooperation. If selection neither favours nor opposes cooperation, then this probability is $1/N$, which is the fixation probability of a neutral mutant. If the fixation

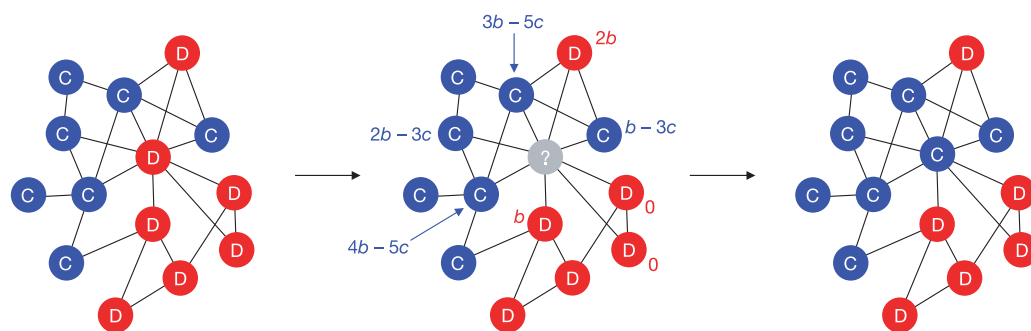


Figure 1 | The rules of the game. Each individual occupies the vertex of a graph and derives a payoff, P , from interactions with adjacent individuals. A cooperator (blue) pays a cost, c , for each neighbour to receive a benefit, b . A defector (red) pays no cost and provides no benefit. The fitness of a player is given by $1 - w + wP$, where w measures the intensity of selection. Strong selection means $w = 1$. Weak selection means $w \ll 1$. For ‘death–birth’

updating, at each time step, a random individual is chosen to die (grey); subsequently the neighbours compete for the empty site in proportion to their fitness. In this example, the central, vacated vertex will change from a defector to a cooperator with a probability $F_C/(F_C + F_D)$, where the total fitness of all adjacent cooperators and defectors is $F_C = 4(1 - w) + (10b - 16c)w$ and $F_D = 4(1 - w) + 3bw$, respectively.

¹Department of Biology, Kyushu University, Fukuoka 812-8581, Japan. ²Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, ³Department of Applied Mathematics, Harvard University, Cambridge, Massachusetts 02138, USA.

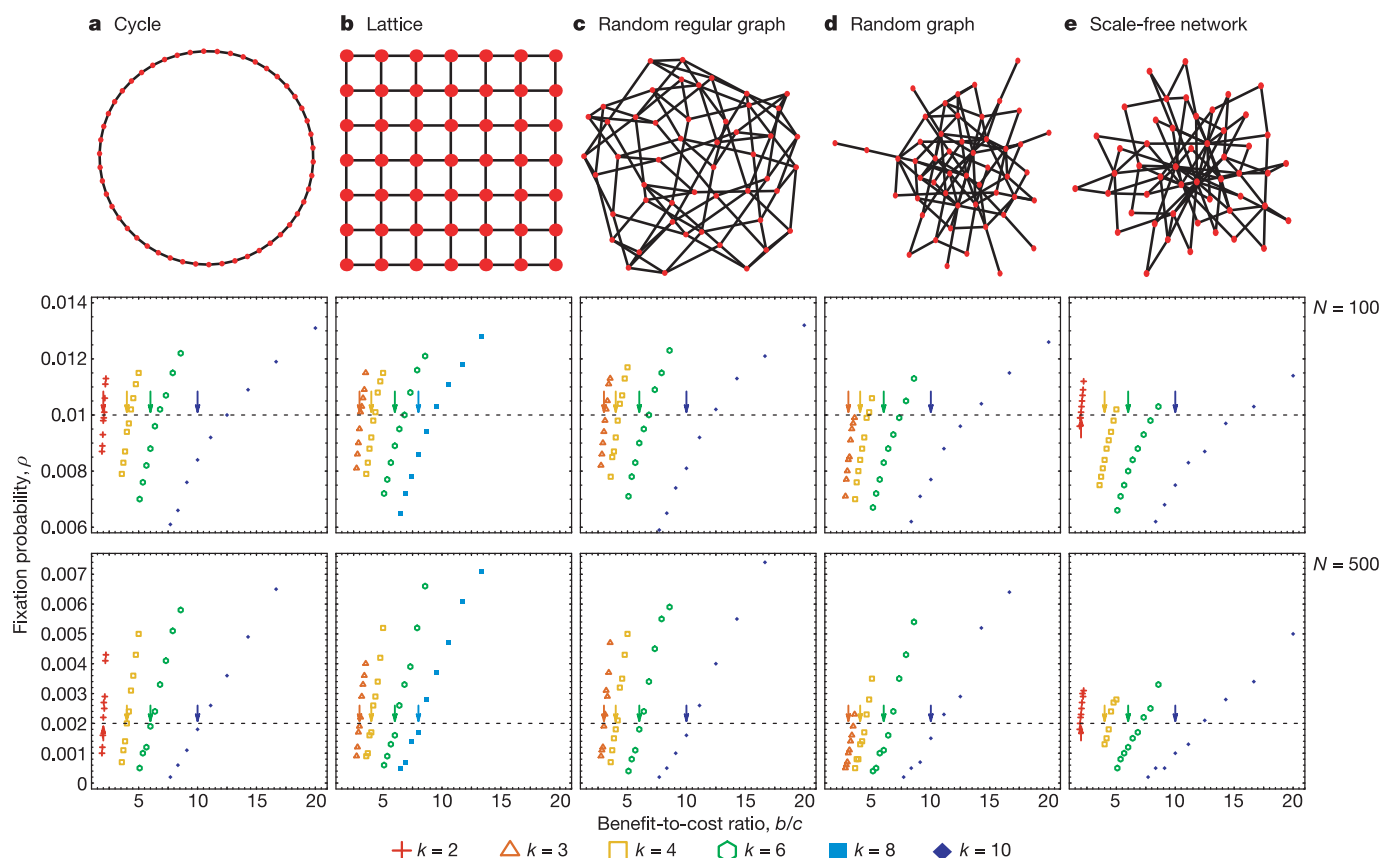


Figure 2 | The simple rule, $b/c > k$, is in good agreement with numerical simulations. The parameter k denotes the degree of the graph, which is given by the (average) number of neighbours per individual. The first row illustrates the type of graph for $k = 2$ (a) and $k = 4$ (b–e). The second and third rows show simulation data for population sizes $N = 100$ and $N = 500$. The fixation probability, ρ , of cooperators is determined by the fraction of runs where cooperators reached fixation out of 10^6 runs under weak selection, $w = 0.01$. Each type of graph is simulated for different (average)

degrees ranging from $k = 2$ to $k = 10$. The arrows mark $b/c = k$. The dotted horizontal line indicates the fixation probability $1/N$ of neutral evolution. The data suggest that $b/c > k$ is necessary but not sufficient. The discrepancy is larger for non-regular graphs (d, e) with high average degree ($k = 10$). This is not surprising given that the derivation of the rule is for regular graphs and in the limit $N \gg k$. Note that the larger population size, $N = 500$, gives better agreement. Interactive online tutorials can be found at <http://univie.ac.at/virtuallabs>.

probability of a single cooperator is greater than $1/N$, then selection favours the emergence of cooperation. We also calculate the fixation probability of a single defector in a population of cooperators, and compare the two fixation probabilities.

The traditional well-mixed population of evolutionary game theory is represented by the complete graph, where all vertices are connected. In this special situation, cooperators are always opposed by selection. This is the fundamental intuition of classical evolutionary game theory. But what happens on other graphs?

Let us first consider a cycle. Each individual is linked to two neighbours. A single cooperator could be wiped out immediately or take over one of its two neighbours. A cluster of two cooperators could expand to three cooperators or revert to a single cooperator. In any case, the lineage starting from one cooperator always forms a single cluster of cooperators, which cannot fragment into pieces. This fact allows a straightforward calculation. We find that selection favours cooperation if $b/c > 2$. This result holds for weak selection and large population size.

Next, we study regular graphs, where each individual has exactly k neighbours. Such graphs include cycles, spatial lattices and random regular graphs. For all such graphs, a direct calculation of the fixation probability is impossible, because a single invader can lead to very complicated patterns: the emerging cluster usually breaks into many pieces, allowing a large number of conceivable geometric configurations. In general, the inherent complexity of games on graphs makes analytical investigations almost always impossible.

Nevertheless, we can calculate the fixation probability of a randomly placed mutant for any two-person, two-strategy game on a regular graph by using pair approximation and diffusion approximation (see Supplementary Information). In particular, we find that cooperators have a fixation probability greater than $1/N$ and defectors have a fixation probability less than $1/N$, if:

$$b/c > k$$

The ratio of benefit to cost of the altruistic act has to exceed the degree, k , which is given by the number of neighbours per individual. This condition is derived for weak selection and under the assumption that the population size, N , is much larger than the degree, k .

We find excellent agreement with numerical simulations (Fig. 2). For a given population size, $b/c > k$ is a necessary condition for selection to favour cooperators. As the population size increases, the discrepancy between $b/c > k$ and the numerical simulations becomes smaller. Moreover, we find that the rule also holds for random graphs²⁵ and scale-free networks^{26,27}, where individuals differ in the number of their neighbours. Here k denotes the average degree of the graph. Scale-free networks fit slightly less well than random graphs, presumably because they have a larger variance of the degree distribution.

The intuitive justification for the $b/c > k$ rule is illustrated in Fig. 3. Consider one cooperator and one defector competing for an empty site. The payoff for the cooperator is $P_C = bq_{C|C}(k-1) - ck$. The payoff for the defector is $P_D = bq_{C|D}(k-1)$. The conditional

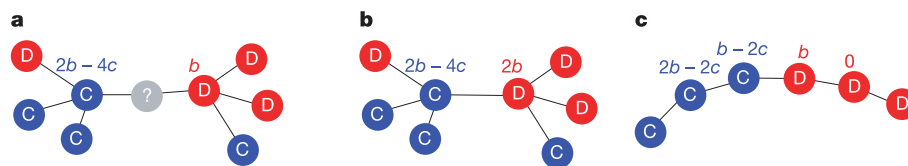


Figure 3 | Some intuition for games on graphs. **a**, For death–birth updating, we must consider a cooperator and a defector competing for an empty site. The pair-approximation calculation shows that for weak selection the cooperator has one more cooperator among its $k - 1$ other neighbours than does the defector. Hence, the cooperator has a higher chance to win the empty site if $b/c > k$. **b**, For birth–death updating, we must consider a cooperator–defector pair competing for the next reproduction event. Again the cooperator has one more cooperator among its $k - 1$ other neighbours than the defector, but the focal cooperator is also a neighbour of the defector.

probability to find a cooperator next to a cooperator is $q_{C|C}$ and to find a cooperator next to a defector is $q_{C|D}$. The cooperator pays cost c for all of its k neighbours and receives benefit b from each cooperator among its $k - 1$ neighbours, excluding the contested site. The defector pays no cost, but receives benefit b from each cooperator among its $k - 1$ neighbours, also excluding the contested site. The payoff that comes from the contested site is excluded, because it contributes equally to the cooperator and the defector and therefore cancels out. If $P_C > P_D$, then selection favours the cooperator. Pair-approximation shows that $(k - 1)(q_{C|C} - q_{C|D}) = 1$ for weak selection. Thus, the cooperator has on average one more cooperator neighbour than the defector. Therefore, we obtain $P_C - P_D = b - ck$, which leads to the $b/c > k$ rule.

We have also explored other update mechanisms. Suppose at each time step a random individual is chosen to update its strategy; it will stay with its own strategy or imitate one of the neighbours proportional to fitness. For this ‘imitation’ updating, we find that cooperators are favoured if $b/c > k + 2$. This result can be obtained with an exact calculation for the cycle and with pair approximation for regular graphs. Again, there is good agreement with numerical simulations (Supplementary Fig. 4). Mathematically, imitation updating can be obtained from our earlier death–birth updating by adding loops to every vertex. Therefore, each individual is also its own neighbour. Let us define the connectivity, k , of a vertex as the total number of links connected to that vertex, noting that a loop is connected twice. Then the simple rule $b/c > k$ holds both for the imitation and death–birth updating.

There are also update rules for which selection can never favour cooperators. For example, let us consider ‘birth–death’ updating: at each time step an individual is selected for reproduction proportional to fitness, and the offspring replaces a randomly chosen neighbour. In this case, selection always favours defectors, because only the payoff of individuals right at the boundary between cooperators and defectors matters, and there cooperators are always at a disadvantage (Fig. 3b). In the two other models, the payoffs of individuals that are one place removed from the boundary also play a role, which gives cooperation a chance to survive.

Using a different model, van Baalen and Rand¹¹ have derived a condition for the initial invasion of cooperators. In their model, the vertices of a spatial lattice (or a graph) are either empty or occupied by cooperators or defectors. There are birth, death and migration events. Implicitly, they have shown that without migration a few cooperators can successfully invade a population of defectors if $b/c > k^2/(k - 1)$. The difference between this result and ours is not surprising. The invasion condition of ref. 11 examines whether rare cooperators are able to increase in abundance, whereas our fixation probability includes the whole evolutionary trajectory including the initial invasion and propagation of cooperators as well as the final extinction of defectors. For a comparison of invasion and fixation criteria see Wild and Taylor³⁰.

Hence, both competitors are linked to the same number of cooperators, and therefore the defector has a higher payoff. For birth–death updating, selection does not favour cooperation. **c**, On a cycle ($k = 2$), the situation is simple. A direct calculation, for weak selection and large population size, leads to the following results. For birth–death updating, the boundary between a cluster of cooperators and defectors tends to move in favour of defectors. For death–birth updating, the cooperator cluster expands if $b/c > 2$. For imitation updating, the cooperator cluster expands if $b/c > 4$.

Thus, we have shown that evolutionary dynamics on graphs can favour cooperation over defection if the benefit to cost ratio, b/c , of the altruistic act exceeds the average connectivity, k . The fewer connections there are, the easier it is for natural selection to promote cooperation. In our present analysis, all connections are equally strong. A next step will be to explore graphs with weighted edges. In social networks, people might have a substantial number of connections, but only very few of them are strong. Hence, the ‘effective’ average degree, k , of many relevant networks could be small, thereby making selection of cooperation on graphs a powerful option.

Our study is theoretically motivated, but has implications for empirical research. For example, one can envisage an experiment where people are asked to play a non-repeated Prisoner’s Dilemma within a given network. Certain network structures should promote cooperative behaviour more than others. In particular, more cooperation should emerge if connectivity is low. Moreover, in certain animal species there exist complicated social networks. Observational studies could reveal how network structure affects the level of cooperation; higher connectivity should reduce cooperation. In this paper, as a logical first step, we have studied the simplest possible interaction between unconditional cooperators and defectors, but in an extended approach, both in terms of theory and experiment, it will be interesting to see which strategies of direct or indirect reciprocity evolve on particular networks.

Finally, we note the beautiful similarity of our finding with Hamilton’s rule¹, which states that kin selection can favour cooperation provided $b/c > 1/r$, where r is the coefficient of genetic relatedness between individuals. The similarity makes sense. In our framework, the average degree of a graph is an inverse measure of social relatedness (or social viscosity). The fewer friends I have the more strongly my fate is bound to theirs.

Received 7 December 2005; accepted 26 January 2006.

1. Hamilton, W. D. The genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 1–16 (1964).
2. Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
3. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
4. Wilson, E. O. *Sociobiology* (Harvard Univ. Press, Cambridge, Massachusetts, 1975).
5. Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **288**, 850–852 (2000).
6. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785–791 (2003).
7. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity. *Nature* **437**, 1291–1298 (2005).
8. Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359**, 826–829 (1992).
9. Killingback, T. & Doebeli, M. Spatial evolutionary game theory: Hawks and Doves revisited. *Proc. R. Soc. Lond. B* **263**, 1135–1144 (1996).
10. Nakamaru, M., Matsuda, H. & Iwasa, Y. The evolution of cooperation in a lattice-structured population. *J. Theor. Biol.* **184**, 65–81 (1997).

11. van Baalen, M. & Rand, D. A. The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* **193**, 631–648 (1998).
12. Mitteldorf, J. & Wilson, D. S. Population viscosity and the evolution of altruism. *J. Theor. Biol.* **204**, 481–496 (2000).
13. Hauert, C., De Monte, S., Hofbauer, J. & Sigmund, K. Volunteering as red queen mechanism for cooperation in public goods games. *Science* **296**, 1129–1132 (2002).
14. Le Galliard, J., Ferriere, R. & Dieckman, U. The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution* **57**, 1–17 (2003).
15. Hauert, C. & Doebeli, M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428**, 643–646 (2004).
16. Ifti, M., Killingback, T. & Doebeli, M. Effects of neighbourhood size and connectivity on the spatial continuous prisoner's dilemma. *J. Theor. Biol.* **231**, 97–106 (2004).
17. Santos, F. C. & Pacheco, J. M. Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.* **95**, 098104 (2005).
18. Levin, S. A. & Paine, R. T. Disturbance, patch formation, and community structure. *Proc. Natl Acad. Sci. USA* **71**, 2744–2747 (1974).
19. Durrett, R. & Levin, S. A. The importance of being discrete (and spatial). *Theor. Popul. Biol.* **46**, 363–394 (1994).
20. Hassell, M. P., Comins, H. N. & May, R. M. Species coexistence and self-organizing spatial dynamics. *Nature* **370**, 290–292 (1994).
21. Skyrms, B. & Pemantle, R. A dynamic model of social network formation. *Proc. Natl Acad. Sci. USA* **97**, 9340–9346 (2000).
22. Abramson, G. & Kuperman, M. Social games in a social network. *Phys. Rev. E* **63**, 030901 (2001).
23. Szabó, G. & Vukov, J. Cooperation for volunteering and partially random partnership. *Phys. Rev. E* **69**, 036107 (2004).
24. Lieberman, E., Hauert, C. & Nowak, M. A. Evolutionary dynamics on graphs. *Nature* **433**, 312–316 (2005).
25. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
26. Barabasi, A. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
27. Taylor, P. D. & Jonker, L. Evolutionary stable strategies and game dynamics. *Math. Biosci.* **40**, 145–156 (1978).
28. Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, UK, 1998).
29. Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).
30. Wild, G. & Taylor, P. D. Fitness and evolutionary stability in game theoretic models of finite populations. *Proc. R. Soc. Lond. B* **271**, 2345–2349 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Support from the John Templeton Foundation, JSPS, NDSEG and Harvard-MIT HST is gratefully acknowledged. The Program for Evolutionary Dynamics at Harvard University is sponsored by Jeffrey Epstein.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.N. (martin_nowak@harvard.edu).

LETTERS

Homology of arthropod anterior appendages revealed by Hox gene expression in a sea spider

Muriel Jager¹, Jérôme Murienne^{1†}, Céline Clabaut^{1†}, Jean Deutsch², Hervé Le Guyader¹ & Michaël Manuel¹

Arthropod head segments offer a paradigm for understanding the diversification of form during evolution, as a variety of morphologically diverse appendages have arisen from them. There has been long-running controversy, however, concerning which head appendages are homologous among arthropods, and from which ancestral arrangement they have been derived. This controversy has recently been rekindled by the proposition that the probable ancestral arrangement, with appendages on the first head segment, has not been lost in all extant arthropods as previously thought, but has been retained in the pycnogonids, or sea spiders¹. This proposal was based on the neuroanatomical analysis of larvae from the sea spider *Anoplodactylus* sp., and suggested that the most anterior pair of appendages, the chelifores, are innervated from the first part of the brain, the protocerebrum. Our examination of Hox gene expression in another sea spider, *Endeis spinosa*, refutes this hypothesis. The anterior boundaries of Hox gene expression domains place the chelifore appendages as clearly belonging to the second head segment, innervated from the second part of the brain, the deutocerebrum. The deutocerebrum must have been secondarily displaced towards the protocerebrum in pycnogonid ancestors. As anterior-most appendages are also deutocerebral in the other two arthropod groups, the Euchelicerata and the Mandibulata, we conclude that the protocerebral appendages have been lost in all extant arthropods.

Fossil evidence suggests that stem-group arthropods possessed a pair of protocerebral appendages, known as the 'frontal appendages' or 'primary antennae', in front of the deutocerebral appendage pair^{2,3}. This idea is also supported by the anatomy of the anterior-most region in onychophorans, a phylum closely related to the arthropods^{4,5}, and from developmental gene expression data suggesting that the rostrum or labrum of arachnids and hexapods includes remnants of this ancestral appendage pair⁶. The existence of protocerebral head appendages in the original arthropod ground plan weakens the classical concept of an 'acron', in which the anterior-most region of the head containing the eyes and protocerebrum was not thought to be segmental.

In recent years, examination of Hox gene expression patterns^{7,8} and immunohistological analysis⁹ have resolved the relationships between cerebral appendages of the two main extant arthropod lineages, the Euchelicerata (arachnids and horseshoe crabs) and the Mandibulata (myriapods, crustaceans and hexapods). These new studies showed clearly that the short prehensile 'chelicerae' of euchelicerates, and the (first) antennae of mandibulates, are homologous, both being innervated by deutocerebral ganglia. It was thought that no extant arthropods retained the ancestral protocerebral appendages in their original form.

Efforts to address this issue have now focused attention on a small group of marine arthropods called the pycnogonids, or sea spiders.

The pycnogonids are positioned phylogenetically either as the sister group of euchelicerates^{10,11} or as the sister group of all remaining arthropods¹². In sea spiders, the first pair of appendages are the chelifores, which are very similar in appearance to the chelicerae of the Euchelicerata. Since the nineteenth century, the majority of zoologists have viewed chelifores and chelicerae as homologous¹³. This homology was recently challenged by Maxmen *et al.*, who provided a detailed neuroanatomical study of the *Anoplodactylus* sp. protonymphon larva¹. The protonymphon larva (Fig. 1a) has three appendage pairs: the chelifores, as well as two other larval pairs of appendages—the first of which corresponds to adult pedipalps and the second to adult ovigers (small legs used by the male to carry the eggs). The novel proposition—that pycnogonid chelifores are protocerebral—was based on the observation that chelifore nerves originate from a ganglion pair closely associated with the dorsal protocerebrum, which innervates the eyes. According to this interpretation, the pycnogonid homologues of chelicerae are not the chelifores, but the second pair of larval appendages (corresponding to the adult pedipalps). This conclusion has major implications for our understanding of arthropod head evolution, because it would make sea-spider chelifores equivalent to the protocerebral appendages proposed for ancestral arthropods^{2,3}.

Given that boundaries of Hox gene expression have been used with success in recent years to clarify homologies between anterior segments among arthropod groups, we looked at Hox gene expression in the protonymphon larva of *Endeis spinosa*, a common sea-spider species on the European Atlantic coast. Routine polymerase chain reaction (PCR) amplification was performed to clone Hox genes relevant to the study of chelifore homology: the anterior Hox genes *labial* (*lab*) and *proboscipedia* (*pb*) (orthologous to *Hox1* and *Hox2*, respectively), and the median Hox gene *Deformed* (*Dfd*) (orthologous to *Hox4*). (Sequence alignments and phylogenetic analyses establishing orthology relationships for these genes are shown in Supplementary Information.) Of special interest is the anterior boundary of pycnogonid *Deformed*, as stated by Budd and Telford¹⁴, because in all investigated arthropods this boundary marks the limit between the tritocerebral segment and the fourth head segment.

In situ hybridizations for the three genes are shown in Fig. 1. The Hox genes *lab* (Fig. 1b) and *pb* (Fig. 1c) are expressed in the ectoderm of segments corresponding to the second and third larval pairs of appendages, but not in the chelifore segment. In contrast, *Dfd* (Fig. 1d–f) is expressed only in the segment bearing the third pair of larval appendages. When compared to Hox gene data in arachnids and mandibulates¹⁵ (Fig. 2), our pycnogonid Hox data unambiguously support the homology of sea-spider chelifores to chelicerae in arachnids and to antennae in mandibulates, and hence the deutocerebral identity of pycnogonid chelifores. Correspondingly, the second pair of larval appendages is homologous to pedipalps in

¹Université Pierre et Marie Curie-Paris 6, UMR 7138 CNRS UPMC MNHN ENS IRD, Case 05, 9 quai St Bernard, 75005 Paris, France. ²Université Pierre et Marie Curie-Paris 6, UMR 7622 CNRS UPMC, 9 quai St Bernard, 75005 Paris, France. [†]Present addresses: Muséum National d'Histoire Naturelle, UMR 5202 CNRS, Département Systématique et Evolution, case 50, 45 rue Buffon, 75005 Paris, France (J.M.); Department of Biology, Evolutionary Biology, University of Konstanz, D-78457 Konstanz, Germany (C.C.).

arachnids and to second antennae (or the intercalary segment) in mandibulates. The third pair of larval appendages (that is, adult ovigers) corresponds to the first walking legs in arachnids and the mandibles in mandibulates.

Consistent with Hox data, published anatomical evidence indicates that in sea spiders, chelifore ganglia and the protocerebrum belong to two distinct head segments. Embryological studies have shown that the paired ganglia innervating the chelifores are formed posterior to the intestine and the protocerebrum, and only later gain a pre-oral position adjacent to the protocerebrum^{16,17}. A second argument concerns the number of neural glands in the pre-oral region. In pycnogonid larvae, each ganglion pair is associated with a pair of neurosecretory glands. The pre-oral region contains two pairs of these glands: one pair associated with chelifore ganglia, and one pair associated with the 'cerebroid' or protocerebral ganglion pair^{18,19}.

At first glance, our results conflict with those of Maxmen *et al.*¹, but they can be reconciled through a re-interpretation of their morphological observations. The location of chelifore ganglia—just in front of, and in close proximity to, the protocerebrum (including optic ganglia and nerves)—is probably a derived character. This implies that pycnogonid ancestors underwent both anterior condensation of the central nervous system, and a rotation of the brain, which brought the chelifore ganglia forward to occupy a position more anterior than the protocerebrum. This derived condition is observed also in arachnids²⁰, with a comparable forward migration of the cheliceral ganglia during embryonic development²¹. In contrast, a straighter brain, with cheliceral ganglia standing in a more posterior location, is found in adult horseshoe crabs⁹, and probably represents the primitive condition in chelicerates. This condensation and rotation process during ontogeny is not unusual for arthropods¹³, and is found convergently: for example, a condensed and rotated brain with the deutocerebrum located in front of the protocerebrum occurs in spiny lobsters among crustaceans²².

Even within the pycnogonid clade, there is a documented tendency towards an anterior concentration of ganglionic structures (cephalization), with this trend culminating in the genus *Anoplodactylus*^{18,23}. Whereas other sea-spider species investigated in the past have a post-oral commissure between ventral ganglia corresponding to the

second larval pair of appendages^{16,18,23}, *Anoplodactylus* sp. larva have a pre-oral commissure in this segment¹. This most likely exemplifies the derived anterior compaction of the nervous system in *Anoplodactylus*, rather than indicating a deutocerebral identity of ganglia targeting the second pair of larval appendages in sea spiders. A way to test this hypothesis would be to investigate the organization of the anterior central nervous system in a wide sample of pycnogonid species, and to reconstruct the evolution of ganglion number and location on a well-supported phylogeny of this group.

Although our data indicate that the protocerebral appendage pair of arthropods, documented in some fossil forms, has not been retained in the form of chelifores in living sea spiders, it may still have left some remnants in the pycnogonid head. Such remnants may be present in the proboscis—a hollow cylindrical anterior structure that shows tri-radiate symmetry, and is used by sea spiders as a sucking device to prey upon soft animals. Some have considered the sea-spider proboscis to be an ancestral structure, homologous to the tri-radiate pharynx located behind the mouth in non-arthropod ecdysozoans such as nematodes^{1,24}. However, an alternative hypothesis, in which the proboscis is a composite structure novel to the pycnogonids, is suggested by the characteristics of its innervation. While the two ventro-lateral thirds (or 'antimeres') of the proboscis are innervated by the ventral ganglia corresponding to the second appendage pair, the dorsal third is innervated by a pair of so-called 'rostral nerves', associated with the cheliforal ganglia, which is reminiscent of rostrum innervation in arachnids^{20,25,26}. Furthermore, the anlage of this dorsal third has a bi-lobate shape^{17,18}, as is the case for the rostrum or labrum of many arthropods^{27,28}. Hence, contrary to the claim by Maxmen *et al.* that "pycnogonids do not possess a labrum"¹, the dorsal third of the proboscis could be the remnant of an additional appendage pair anterior to the chelifores, similar to the rostrum/labrum in other arthropods.

Whether or not this is the case, our data imply that protocerebral appendages were lost or dramatically reduced in a common ancestor of extant arthropods. Furthermore, our results consolidate the grounds for homology between chelifores/chelicerae and the (first) antennae of mandibulates, implying that all of the anterior-most appendages of extant arthropods are homologous and are

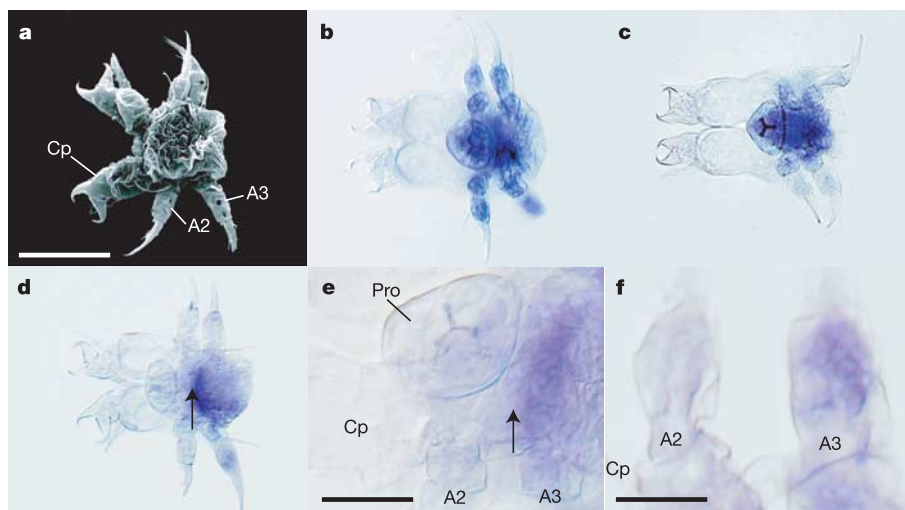


Figure 1 | Hox gene expression in newly hatched *Endeis spinosa* protonymphon larvae. Anterior is to the left in all pictures. **a**, Scanning electron microscopy (SEM) of the protonymphon larva, showing its external morphology in dorsal view. Cp, chelifore; A2, second larval appendage; A3, third larval appendage. **b**, Expression of *labial*. The signal is located in the ectoderm of the second and third larval appendage pairs, and in the corresponding trunk region. **c**, Expression of *proboscipedia* occurs in the same body region as the expression of *labial*, except that in the legs (larval appendage pairs 2 and 3) the signal is restricted to the basal article.

d–f, Expression of *Deformed* (in three different larvae). Arrows in **d** and **e** indicate the anterior boundary of *Deformed* expression in the trunk. Panel **e** shows the trunk and the base of the appendages (labels as in **a**) at a higher magnification. The round structure is the larval proboscis (Pro). A detailed view of the legs is provided in panel **f**. Although there is a strong epidermal staining in the third appendage pair, the second appendage pair shows no staining. A probe targeting a fourth Hox gene, *Sex comb reduced* (*Scr*) (orthologous to *Hox5*), produced no staining. Scale bars, 50 μ m (**a**), 17 μ m (**e**), 12 μ m (**f**); magnification in **b–d** is as in **a**.

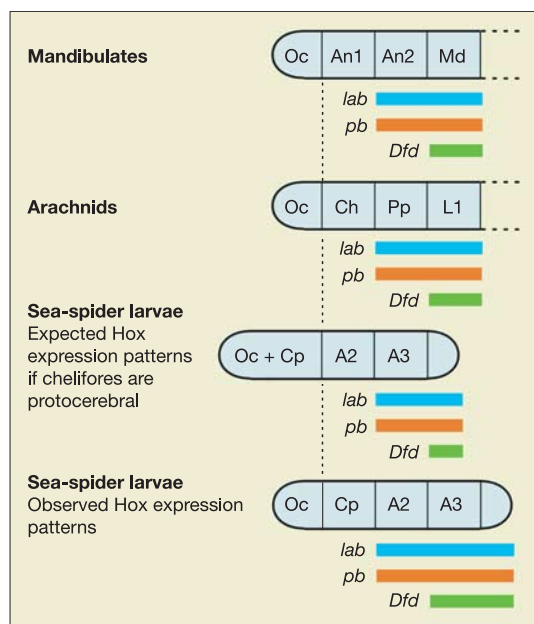


Figure 2 | Expression patterns of *lab*, *pb* and *Dfd*, and inferred correspondence among anterior body segments in mandibulates, arachnids and the protonymphon larva of pycnogonids. Data for Hox gene expression in mandibulates and arachnids are from ref. 15. The dashed vertical line indicates the limit between the protocerebral and deutocerebral segments. A2, second pair of larval appendages; A3, third pair of larval appendages; An1, first antennae; An2, second antennae; Ch, chelicerae; Cp, chelifores; L1, first pair of walking legs; Md, mandibles; Oc, ocular segment; Pp, pedipalps.

deutocerebral. In addition, although the protocerebral chelifore hypothesis favoured a basal position of pycnogonids with respect to other arthropods, the homology between chelifores and chelicerae is more consistent with a close phylogenetic relationship between pycnogonids and euchelicerates^{10,11}.

METHODS

Specimen collection. Adult specimens of *E. spinosa* were collected in Roscoff (France), during their reproductive period in June. Adult males carrying egg balls on their ovigers were kept for a few days in filtered natural sea water until hatching of the larvae.

Hox gene cloning and expression analysis. Partial Hox sequences were amplified by standard PCR methods using degenerate oligonucleotide primers corresponding to amino-acid sequences ELEKEF (5'-GARYTRGARAARG ARTT-3') and WFQNR (5'-YCKYCKRTTYTGRAACCA-3'). Genomic DNA and larval and adult complementary DNA were used as templates in independent experiments. PCR products (117 base pairs) were cloned and sequenced. Sequences were then extended by 3' - and 5' - RACE (rapid amplification of cDNA ends) PCR, using, for each gene, two sets of specific (nondegenerate) primers designed from the initial partial sequence identified during the preceding step. For *in situ* hybridization, newly hatched larvae were fixed overnight at 4°C in 4% paraformaldehyde in PBS. To alter the larval cuticle, they were incubated for three days at 4°C in 5% β-mercaptoethanol in 0.1 M Tris-HCl, pH 7.5, 1% Triton X-100. Otherwise, we followed standard procedures for whole-mount *in situ* hybridization using digoxigenin-labelled antisense riboprobes, as explained in ref. 29. To avoid cross-hybridization, probes were transcribed from variable regions outside and upstream of the homeodomain. Their lengths were 508, 480 and 540 bp, respectively, for *lab*, *pb* and *Dfd*. Negative controls (with a sense probe) gave no staining.

Received 4 November 2005; accepted 18 January 2006.

- Maxmen, A., Browne, W. E., Martindale, M. Q. & Giribet, G. Neuroanatomy of sea spiders implies an appendicular origin of the protocerebral segment. *Nature* **437**, 1144–1148 (2005).
- Budd, G. E. A palaeontological solution to the arthropod head problem. *Nature* **417**, 271–275 (2002).
- Scholtz, G. & Edgecombe, G. Heads, Hox and the phylogenetic position of trilobites. *Crustac. Issues* **16**, 139–165 (2005).
- Eriksson, B. J., Tait, N. N. & Budd, G. E. Head development in the onychophoran

Euperipatoides kanangrensis with particular reference to the central nervous system. *J. Morphol.* **255**, 1–23 (2003).

- Mayer, G. & Koch, M. Ultrastructure and fate of the nephridial anlagen in the antennal segment of *Epiperipatus biolleyi* (Onychophora, Peripatidae)—evidence for the onychophoran antennae being modified legs. *Arthrop. Struct. Dev.* **34**, 471–480 (2005).
- Popadic, A., Panganiban, G., Rusch, D., Shear, W. A. & Kaufman, T. C. Molecular evidence for the gnathobasic derivation of arthropod mandibles and for the appendicular origin of the labrum and other structures. *Dev. Genes Evol.* **208**, 142–150 (1998).
- Damen, W. G. M., Hausdorf, M., Seyfarth, E.-A. & Tautz, D. A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. *Proc. Natl Acad. Sci. USA* **95**, 10665–10670 (1998).
- Telford, M. J. & Thomas, R. H. Expression of homeobox genes shows chelicerate arthropods retain their deutocerebral segment. *Proc. Natl Acad. Sci. USA* **95**, 10671–10675 (1998).
- Mittmann, B. & Scholtz, G. Development of the nervous system in the “head” of *Limulus polyphemus* (Chelicerata: Xiphosura): morphological evidence for a correspondence between the segments of the chelicerae and of the (first) antennae of Mandibulata. *Dev. Genes Evol.* **213**, 9–17 (2003).
- Mallatt, J. M., Garey, J. R. & Schultz, J. W. Ecdysozoan phylogeny and bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. *Mol. Phyl. Evol.* **31**, 178–191 (2004).
- Dunlop, J. A. & Arango, C. P. Pycnogonid affinities: a review. *J. Zool. Syst. Evol. Res.* **43**, 8–21 (2005).
- Giribet, G., Edgecombe, G. D. & Wheeler, W. C. Arthropod phylogeny based on eight molecular loci and morphology. *Nature* **413**, 157–161 (2001).
- Brusca, R. C. & Brusca, G. J. *Invertebrates* 2nd edn (Sinauer, Sunderland, Massachusetts, 2003).
- Budd, G. E. & Telford, M. J. Along came a sea spider. *Nature* **437**, 1099–1102 (2005).
- Hughes, C. L. & Kaufman, T. C. Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* **4**, 459–499 (2002).
- Meisenheimer, J. Beiträge zur entwicklungsgeschichte der pantopoden. I. Die entwicklung von *Ammothea echinata* hodge bis zur ausbildung der larvenform. *Z. Wiss. Zool.* **72**, 191–248 (1902).
- Winter, G. Beiträge zur morphologie und embryologie des vorderen körperabschnitts (Cephalosoma) der Pantopoda gerstaecker, 1863. *Z. Zool. Syst. Evol.-Forsch.* **18**, 27–61 (1980).
- Sanchez, S. *Le Développement des Pycnogonides et leurs Affinités avec les Arachnides*. Thesis, CNRS, Paris (1959).
- Morgan, T. H. A contribution to the embryology and phylogeny of the pycnogonids. *Stud. Biol. Lab. Johns Hopkins Univ.* **5**, 1–76 (1891).
- Babu, K. S. Anatomy of the central nervous system of arachnids. *Zool. Jb. Anat.* **82**, 1–154 (1965).
- Weygoldt, P. in *Neurobiology of Arachnids* (ed. Barth, F. G.) 20–37 (Springer, Berlin/Heidelberg/New York, 1985).
- Sandeman, D. C., Scholtz, G. & Sandeman, R. Brain evolution in decapod Crustacea. *J. Exp. Zool.* **295**, 112–133 (1993).
- Henry, L. M. The nervous system of the pycnogonids. *Microentomology USA* **18**, 16–36 (1953).
- Schmidt-Rhaesa, A., Bartolomeus, T., Lemburg, C., Ehlers, U. & Garey, J. R. The position of the Arthropoda in the phylogenetic system. *J. Morphol.* **238**, 263–285 (1998).
- Wiren, E. Zur morphologie und phylogenie der pantopoden. *Zool. Bidr. Uppsala* **6**, 41–181 (1918).
- Bullock, T. H. & Horridge, G. A. *Structure and Function in the Nervous Systems of Invertebrates* Vol. 2 (Freeman, San Francisco/London, 1965).
- Scholtz, G. in *Arthropod Relationships* (eds Fortey, R. A. & Thomas, R. H.) 317–332 (Chapman and Hall, London, 1997).
- Simonnet, F., Deutsch, J. & Quéinnec, E. *hedgehog* is a segment polarity gene in a crustacean and a chelicerate. *Dev. Genes Evol.* **214**, 537–545 (2004).
- Gibert, J. M., Mouchel-Vielh, E., Quéinnec, E. & Deutsch, J. S. Barnacle duplicate *engrailed* genes: divergent expression patterns and evidence for a vestigial abdomen. *Evol. Dev.* **2**, 194–202 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Station Biologique de Roscoff for providing laboratory facilities for specimen collection and preparation. We are grateful to E. Quéinnec, N. Rabet and P. Bunje for advice and discussion, to P. Lamarre for technical help and to T. Jaffredo for laboratory facilities. E. Houliston and G. Scholtz provided much help and insight. This work was funded by the CNRS and the French Ministry of Research.

Author Information Sequences from this work have been deposited in the GenBank database with the following accession numbers: DQ315728 (*E. spinosa lab*); DQ315730 (*E. spinosa pb*); DQ315733 (*E. spinosa Dfd*); and DQ315734 (*E. spinosa Scr*). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M. (Michael.Manuel@snv.jussieu.fr).

Somatic stem cell niche tropism in *Wolbachia*

Horacio M. Frydman^{1,2}, Jennifer M. Li^{1,2}, Drew N. Robson² & Eric Wieschaus^{1,2}

Wolbachia are intracellular bacteria found in the reproductive tissue of all major groups of arthropods^{1,2}. They are transmitted vertically from the female hosts to their offspring, in a pattern analogous to mitochondria inheritance. But *Wolbachia* phylogeny does not parallel that of the host, indicating that horizontal infectious transmission must also occur^{3–5}. Insect parasitoids are considered the most likely vectors, but the mechanism for horizontal transfer is largely unknown^{4,6,7}. Here we show that newly introduced *Wolbachia* cross several tissues and infect the germline of the adult *Drosophila melanogaster* female. Through investigation of bacterial migration patterns during the course of infection, we found that *Wolbachia* reach the germline through the somatic stem cell niche in the *D. melanogaster* germarium. In addition, our

data suggest that *Wolbachia* are highly abundant in the somatic stem cell niche of long-term infected hosts, implying that this location may also contribute to efficient vertical transmission. This is, to our knowledge, the first report of an intracellular parasite displaying tropism for a stem cell niche.

The presence of *Wolbachia* in 20–80% of all insect species^{1,2} is attributed to extensive horizontal transfer. However, propagation within a new population ultimately requires vertical transmission, which entails not only infection of the host soma, but also infection of its germline. Previous studies using a number of different insect and crustacean species have shown that inoculation of *Wolbachia* into the abdominal cavity of females results in stable vertically transmitted infection^{8–10}, suggesting that *Wolbachia* is able to reach

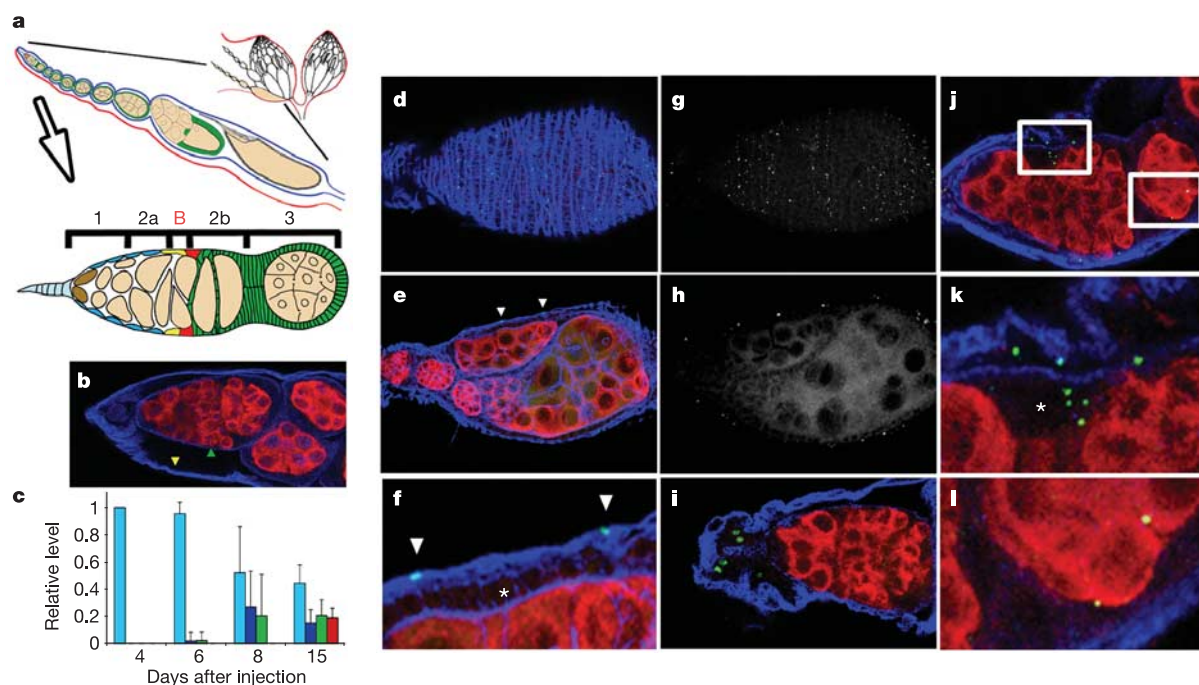


Figure 1 | *Wolbachia* introduced in the abdominal cavity reaches the germline. **a**, A drawing of the *Drosophila* ovary (top right) surrounded by a thin peritoneal sheath (red). Each ovariole (middle) is encased by a muscle epithelium (blue). Egg chambers are formed in the germarium (magnified at the bottom). Germarial regions are indicated. As the germline progresses from region 2a to 2b, it becomes enveloped by the somatic cells (green) derived from 2–3 somatic stem cells (SSC, red cells). For the purpose of this study, region 2 is divided in 2a, ‘border’ and 2b. The ‘border’ contains the SSC and its niche (*B, in red), formed by the anteriorly localized inner germarium sheath cell (IGS, yellow cells) and extracellular matrix. **b**, Control germarium injected with haemolymph from non-infected flies. Germline is in red (all figures). Laminin at muscle (yellow arrow) and germarium basement membrane (green arrow) are shown in blue. **c**, Relative levels of *Wolbachia* at muscle (light blue), muscle lumen (dark

blue), follicle cells (green) and germline (red), at different days after *Wolbachia* injection ($n = 33$ ovarioles; 2,392 bacteria particles; error bars represent standard deviation). **d–l**, Ovaries of hosts injected with *Wolbachia* (green). Germline is shown in red. In **d–f**, muscle and follicle cell cortex actin are in blue. In **i–l**, laminin at muscle and germarium basement membrane are shown in blue. **d**, Ovary muscle 6 days after injection, showing muscle actin fibres with high levels of *Wolbachia*. **g**, *Wolbachia* channel. **e**, Same ovary as in **d**, germline plane. *Wolbachia* (arrowheads) at the muscle. **h**, *Wolbachia* channel of **e**. **f**, Higher magnification of **e**. *Wolbachia* not present at the follicular epithelium (asterisk). **i**, Germarium 8 days after injection. *Wolbachia* present at the muscle lumen around terminal filament. **j**, Germarium 15 days after injection. Boxes are regions depicted at higher magnification in **k** and **l**. *Wolbachia* present in the border region (asterisk in **k**) and germline (**l**).

¹Howard Hughes Medical Institute, ²Department of Molecular Biology, Princeton University, Princeton, New Jersey 08540, USA.

the germline from the infected soma. However, the mechanism underlying this mode of infection has not been characterized. To monitor the dynamics of germline colonization, we injected the *Wolbachia* wMel strain into the abdominal cavity of uninfected *D. melanogaster* females (see Methods) and imaged ovaries at different days (see below). Stocks established from injected flies still maintained *Wolbachia* in the germline 6, 20 and 29 generations after initial injection (Supplementary Fig. 1), indicating that our injection resulted in invasion of the germline, production of mature eggs containing *Wolbachia* and ultimately stable vertical transmission.

To reach the female host germline, injected bacteria need to cross three different tissues: a peritoneal sheath membrane surrounding the entire ovary, a muscle epithelium enclosing each individual ovariole and the somatic tissue enveloping the germline (Fig. 1a). We find that *Wolbachia* sequentially infect tissues of the host ovary in discrete stages (Fig. 1c) and ultimately colonize the germline 15 days following inoculation (Fig. 1l). Furthermore, the pattern of infection suggests that the route to reach the germ cells is through the germarium (Fig. 1). By days 4 and 6, high levels of *Wolbachia* were seen at the muscle surface (Fig. 1d–h), and between the muscle and the germarium basement membrane after 8 days (Fig. 1i). 65% (117 out of 220) of the *Wolbachia* associated with the germaria at this stage

were found in the lumen between the terminal filament and the overlying muscle layer. However, *Wolbachia* entered the ovary preferentially at the border of the germarial regions 2a and 2b (hereafter referred as the 'border' region, Fig. 1j, k). This site is notable for harbouring two to three somatic stem cells (SSC)^{11,12} (Fig. 1a).

The injection experiments suggest that specific somatic regions of the germaria may provide preferred sites for targeting during new infection. To determine whether the same regions would also be targeted in the case of long term maternally transmitted infections, we examined the *Wolbachia* levels of specific ovarian cell types in long term maternally infected flies. In such females, the germline cytoplasm contains higher levels of *Wolbachia* than the cytoplasm of follicle cells at all stages of oogenesis (Supplementary Fig. 2a). The only exception was in the border region, where we observe a high accumulation of *Wolbachia* (Fig. 2a–c) reminiscent of transient infections (Fig. 2i).

To determine whether the accumulation of *Wolbachia* in the border region is mainly concentrated in the SSC themselves, we characterized the *Wolbachia* distribution in germaria in which the SSC were marked by labelling of clonal lineages. These experiments indicated that the highest bacterial levels are observed anterior to the stem cells, often distributed in two major clusters (Fig. 2d–f). In 8 out

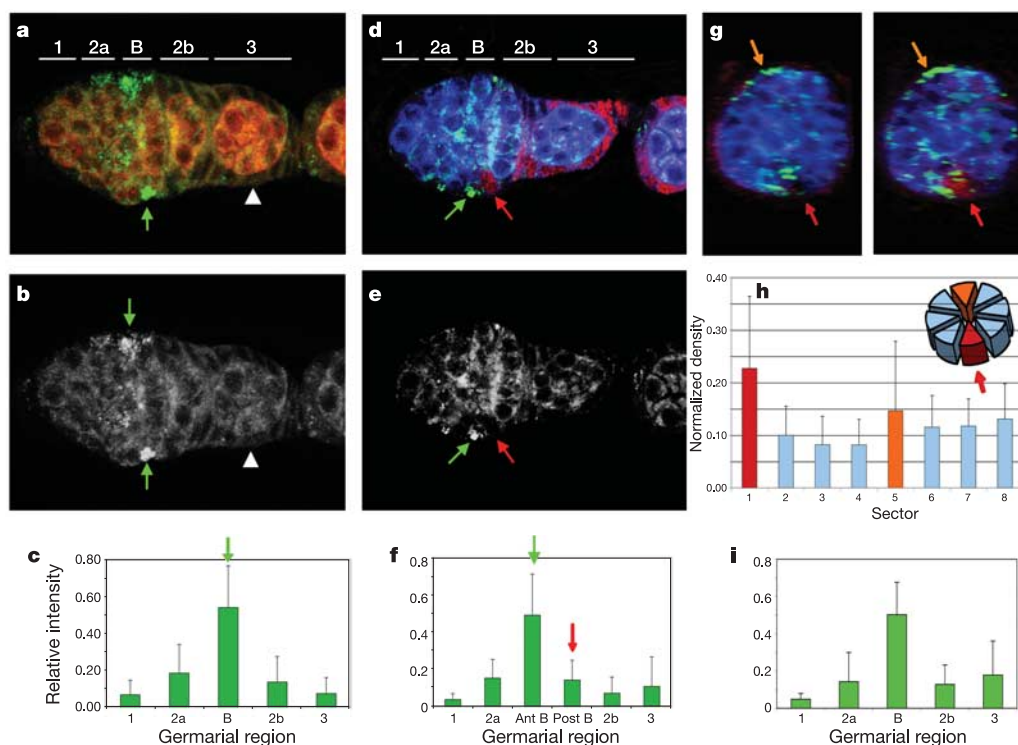


Figure 2 | *Wolbachia* preferentially infect the somatic stem cell niche (SSCN). **a**, Germarium of a maternally infected stock. High levels of *Wolbachia* (green) are seen in the somatic tissue at the 2a/2b border (regions indicated on top). Most of the remaining somatic tissue (arrowhead) has low levels of *Wolbachia* relative to the germline (red). **b**, *Wolbachia* channel, showing two major clusters of *Wolbachia*. **c**, Relative intensities of *Wolbachia* in the soma at the different germarial regions ($n = 26$). Green arrows in **a**, **b** and **c** point to *Wolbachia* accumulation at the border. **d**, SSC and its lineage are labelled in red, *Wolbachia* in green, germline in blue. SSC is the most anteriorly labelled cell (see Methods). *Wolbachia* accumulation (green arrow) is at the SSCN, anterior to the SSC (red arrow). **e**, *Wolbachia* channel from **d**. **f**, The border region is divided into the anterior portion containing the SSCN (green arrow) and the posterior portion containing the SSC (red arrow). Relative *Wolbachia* intensities are shown ($n = 14$). **g**, Overlays of reconstructed slice data at the SSC (red staining on right image) and immediately anterior to the SSC (left image). Reconstructions were

performed to yield slice data perpendicular to the anterior-posterior axis. A radial axis consisting of eight sectors was defined on the SSC plane and oriented so that sector 1 contained the SSC (red arrow). This radial axis was then used to divide the slices of the 2a/2b border region that were anterior to the SSC into eight sectors (see drawing at **h**). Arrows point to *Wolbachia* staining at the sector defined by the SSC (red arrow) and at the sector radially opposed to it (orange arrow, see below). **h**, Normalized *Wolbachia* density by sector in the portion of the 2a/2b border region anterior to the SSC. Sector 1 (red bar at histogram and red arrow at drawing) contains the SSCN which is anterior to the labelled SSC. As the 2a/2b border region usually contains two radially opposed SSC as well as their associated SSCN, sector 5 (orange) may contain the SSCN of a second unmarked SSC ($n = 10$). **i**, Relative intensity of *Wolbachia* in ovarioles from uninfected flies inoculated with *Wolbachia*. Relative levels in the soma for germarial regions 15 days after injection are similar to maternally infected flies ($n = 8$). Error bars in **c**, **f**, **h**, **i** represent standard deviation.

of 10 cases, the highest levels of *Wolbachia* were contained within a sector directly anterior to the SSC ($P = 0.037$; Fig. 2g, h). This sector is predominantly composed of inner germarium sheath (IGS) cells anterior to the SSC that are constituents of the microenvironment harbouring the SSC, the so-called somatic stem cell niche (SSCN, Fig. 1a)^{13–15}.

Could the infected SSCN itself be a source of *Wolbachia* into the germline? Germline cells contact the SSCN as they pass through this region, and afterwards remain in contact with precursors of the follicle cell epithelium arising from the SSC. Thus, an infected niche can potentially transmit *Wolbachia* to the germline by two routes, either directly or by infection of the SSC and its derived follicle cell population. We have evidence suggesting both of these possibilities. In occasional germaria with only one infected niche, we can compare levels of *Wolbachia* in the half containing the infected SSCN to those in the other half. In 16 such germaria (Supplementary Fig. 2b–d and Supplementary Movie), somatic and germline regions anterior to the SSCN show no bias in *Wolbachia* levels in the two halves of the germarium (Supplementary Table). However, the somatic regions posterior to the SSCN show increased *Wolbachia* levels in the half containing the infected SSCN in 14 out of 16 cases, with an average increase of $62 \pm 21\%$, suggesting that a SSC residing in an infected SSCN also becomes infected and its follicular progeny consistently have a greater degree of *Wolbachia* infection ($P = 1.6 \times 10^{-5}$). In the border region, we also see an increase of $52 \pm 22\%$ in germline *Wolbachia* levels in the half of the germarium containing the infected niche ($P = 1.9 \times 10^{-3}$). However, that bias does not seem to be maintained as the interconnected cyst of germ cells continues its growth.

Although *Wolbachia* are found primarily in the reproductive

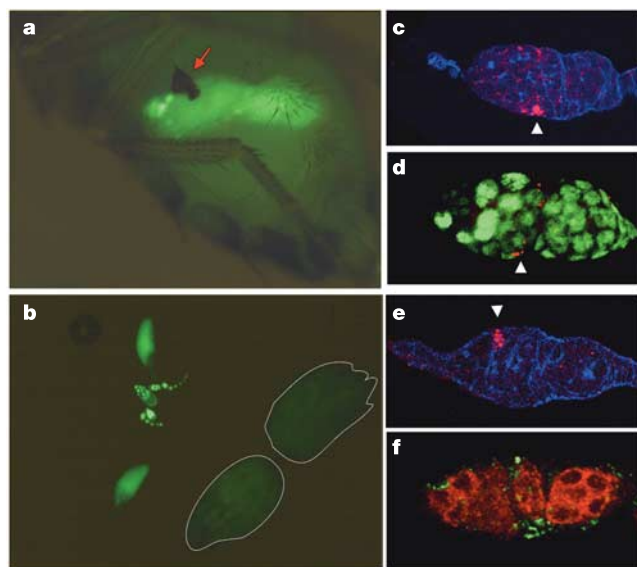


Figure 3 | *Wolbachia* at the abdominal cavity of maternally infected flies infect newly introduced germarium. **a**, Two germaria from a histone-GFP tetracycline treated stock (*Wolbachia*-free) transplanted into the abdominal cavity of a maternally infected fly (see Methods). The dark spot (arrow) is a melanotic scar from transplantation surgery. **b**, Fly from **a** dissected. Both transplanted germaria developed into GFP labelled ovarioles. The host germarium is outlined. **c**, Host germarium from **b** has high levels of *Wolbachia* (red) in the border region (arrowhead). Cortical actin is shown in blue. **d**, *Wolbachia* (red) infects the SSCN of transplanted histone-GFP (green) germarium. Frequently *Wolbachia* levels are considerably lower in the transplanted germarium than in host germaria (arrowhead). **e**, Transplanted germarium from **b**, with moderate levels of *Wolbachia* (red) in one of the SSCN (arrowhead). Actin is shown in blue. **f**, Germarium from *D. mauritiana*. High levels of *Wolbachia* wMau (green) are present in the border region. The germline is shown in red.

tissues, they can also be detected throughout the soma of many insects^{16–19}. We wondered whether *Wolbachia* present in the abdominal cavity of long term maternally infected hosts could serve as an additional source of bacteria into the germline, possibly via infection and accumulation in the SSCN. To test this hypothesis, we transplanted germaria into the abdominal cavity of infected *D. melanogaster* hosts (see Methods). If the vertically transmitted *Wolbachia* present in the abdominal cavity are competent to infect the germline and the SSCN, we should expect to see infection in the transplanted germarium. Two to three weeks after germarium transplantation, the SSCN had been colonized in 22 of 29 cases (Fig. 3a–e). *Wolbachia* levels were higher than in neighbouring somatic tissue in 18 out of the 22 cases. Five of the seven transplanted germaria with no *Wolbachia* in the niche had no *Wolbachia* in the germline, whereas 22 out of 22 germaria with infected niches had *Wolbachia* in the germline. These results demonstrate that endogenous levels of free *Wolbachia* present in the abdomen of maternally infected flies are sufficient to cause infection of germ cells that were previously free of the bacteria. We conclude that the likely source of bacteria into the germline is the infection of the SSCN.

The above experiment has shown that tropism for the stem cell niche and ultimate transfer to germ cells is not merely a property of newly injected bacteria, but may represent a mechanism that contributes to germline infection in long term maternal transmission. Re-infection of the germline through the SSCN might provide a route for *Wolbachia* to maintain infection in species like *Drosophila mauritiana* where, unlike *D. melanogaster*, no special accumulation is observed in germline precursors during early embryonic development^{20,21}. We examined *Wolbachia* distributions in females from a maternally infected *D. mauritiana* line to determine whether if such a pathway is feasible. In the *D. mauritiana* ovary, we detect limited *Wolbachia* levels in most cell types but, consistent with the tropism presented above in *D. melanogaster*, we see high levels in the border region. Levels in the germline are lower, consistent with the possibility that enrichment in the SSCN might provide a stable reservoir for infecting the adjacent germ cells (Fig. 3f). The same enrichment at the border was also observed in *Wolbachia* wRi infecting *Drosophila simulans* and *Wolbachia* wPop infecting *D. melanogaster* (Supplementary Fig. 2e, f).

Both the somatic and germline stem cells require high mitotic rates in order to sustain the high rate of egg production. Under conditions of low or fluctuating *Wolbachia* levels, these proliferation rates may lead to occasional daughter cells with *Wolbachia* levels insufficient for efficient transmission. By contrast, the stromal neighbours that form the niche for these stem cells are more stable and undergo few if any divisions. We propose that the *Wolbachia* infection we observe in such stable niches contributes to the efficient spread of *Wolbachia* within a new host population by providing a point for stable accumulation before subsequent vertical transfer into stem cell derivatives. After introduction of *Wolbachia* into a new host organism, such as following a wasp or parasite mediated infection^{4,6,7}, the limited number of bacteria may be under selective pressure to target the germ cells. Under such low density conditions, direct entry into the germline would result in a reduced number of infected eggs. However, initial targeting of the niche offers the opportunity of population self-renewal, amplification and spreading into the germline, in a strategy similar to that of the SSC themselves. In the case of strict vertical transmission, if the levels of *Wolbachia* in the germline stem cells are not enough to supply every derived germline cyst with *Wolbachia*, the constant and renewed infection of the niche would provide an additional reservoir of *Wolbachia* for entry into the germline. The pathway we describe for infecting the SSCN is therefore a very efficient strategy for both horizontal and vertical transmission.

The mechanism for *Wolbachia* entry and accumulation at the niche is unknown. It has recently been reported that IGS cells originate near the tip of the ovary, where they encase the developing

cyst and move towards the border, turning over at this region²². The dynamic behaviour of IGS cells may contribute to the accumulation of *Wolbachia* at the niche. We anticipate that future elucidation of the cellular basis for horizontal transfer and the mechanism responsible for *Wolbachia* tropism towards the SSCN will deepen our understanding of *Wolbachia* transmission and of the niche maintaining SSC.

METHODS

Immunocytochemistry. Ovarioles were stained as previously described^{23,24}. The following antisera were used at the indicated dilutions: anti-hsp60^{24,25} (Sigma) (1:50), rabbit polyclonal anti-Laminin A (a gift from H. Gutzeit; 1:3,000), rabbit polyclonal anti- β -galactosidase antibody (Cappel) (1:3,000), rat anti-vasa (a gift from P. Lasko; 1:500). Actin was stained with Alexa568 (Molecular Probes) conjugated to phalloidin (1:400).

Injections and transplantations. *Wolbachia* were isolated from adult flies as described previously⁹, with modifications. To ensure the same initial inoculation, all flies analysed in Fig. 1 were injected with 0.2 μ l of the same preparation. Germarium transplantations were performed as described²⁶ using uninfected 1–3-day-old histone-GFP flies²⁷ as germarium donors into newly eclosed (1–6 h) infected flies as hosts.

Received 28 November 2005; accepted 30 March 2006.

- Werren, J. H. & Windsor, D. M. *Wolbachia* infection frequencies in insects: evidence of a global equilibrium? *Proc. R. Soc. Lond. B* **267**, 1277–1285 (2000).
- Jeyaprasath, A. & Hoy, M. A. Long PCR improves *Wolbachia* DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect Mol. Biol.* **9**, 393–405 (2000).
- Werren, J. H. & Bartos, J. D. Recombination in *Wolbachia*. *Curr. Biol.* **11**, 431–435 (2001).
- Vavre, F., Fleury, F., Lepetit, D., Fouillet, P. & Bouletreau, M. Phylogenetic evidence for horizontal transmission of *Wolbachia* in host-parasitoid associations. *Mol. Biol. Evol.* **16**, 1711–1723 (1999).
- O'Neill, S. L., Giordano, R., Colbert, A. M., Karr, T. L. & Robertson, H. M. 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proc. Natl Acad. Sci. USA* **89**, 2699–2702 (1992).
- Huigens, M. E., de Almeida, R. P., Boons, P. A., Luck, R. F. & Stouthamer, R. Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing *Wolbachia* in *Trichogramma* wasps. *Proc. R. Soc. Lond. B* **271**, 509–515 (2004).
- Heath, B. D., Butcher, R. D., Whitfield, W. G. & Hubbard, S. F. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Curr. Biol.* **9**, 313–316 (1999).
- Grenier, S. et al. Successful horizontal transfer of *Wolbachia* symbionts between *Trichogramma* wasps. *Proc. R. Soc. Lond. B* **265**, 1441–1445 (1998).
- Kang, L. et al. Superinfection of *Laodelphax striatellus* with *Wolbachia* from *Drosophila simulans*. *Heredity* **90**, 71–76 (2003).
- Rigaud, T., Pennings, P. S. & Juchault, P. *Wolbachia* bacteria effects after experimental interspecific transfers in terrestrial isopods. *J. Invertebr. Pathol.* **77**, 251–257 (2001).
- Zhang, Y. & Kalderon, D. Hedgehog acts as a somatic stem cell factor in the *Drosophila* ovary. *Nature* **410**, 599–604 (2001).
- Margolis, J. & Spradling, A. Identification and behaviour of epithelial stem cells in the *Drosophila* ovary. *Development* **121**, 3797–3807 (1995).
- Fuchs, E., Tumber, T. & Guasch, G. Socializing with the neighbors: stem cells and their niche. *Cell* **116**, 769–778 (2004).
- Ohlstein, B., Kai, T., Decotto, E. & Spradling, A. The stem cell niche: theme and variations. *Curr. Opin. Cell Biol.* **16**, 693–699 (2004).
- Li, L. & Xie, T. Stem cell niche: structure and function. *Annu. Rev. Cell. Dev. Biol.* **21**, 605–631 (2005).
- Dobson, S. L. et al. *Wolbachia* infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biol.* **29**, 153–160 (1999).
- Min, K. T. & Benzer, S. *Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *Proc. Natl Acad. Sci. USA* **94**, 10792–10796 (1997).
- Cheng, Q. et al. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. *Med. Vet. Entomol.* **14**, 44–50 (2000).
- Mcgraw, E. A. & O'Neill, S. L. *Wolbachia pipiensis*: intracellular infection and pathogenesis in *Drosophila*. *Curr. Opin. Microbiol.* **7**, 67–70 (2004).
- Hadfield, S. J. & Axton, J. M. Germ cells colonized by endosymbiotic bacteria. *Nature* **402**, 482 (1999).
- Veneti, Z., Clark, M. E., Karr, T. L., Savakis, C. & Bourtzis, K. Heads or tails: host-parasite interactions in the *Drosophila*-*Wolbachia* system. *Appl. Environ. Microbiol.* **70**, 5366–5372 (2004).
- Decotto, E. & Spradling, A. C. The *Drosophila* ovarian and testis stem cell niches: similar somatic stem cells and signals. *Dev. Cell* **9**, 501–510 (2005).
- Frydman, H. M. & Spradling, A. C. The receptor-like tyrosine phosphatase lar is required for epithelial planar polarity and for axis determination within *Drosophila* ovarian follicles. *Development* **128**, 3209–3220 (2001).
- Ferree, P. M. et al. *Wolbachia* utilizes host microtubules and dynein for anterior localization in the *Drosophila* oocyte. *PLoS Pathogens* **1**, e14 (2005).
- Mcgraw, E. A., Merritt, D. J., Droller, J. N. & O'Neill, S. L. *Wolbachia* density and virulence attenuation after transfer into a novel host. *Proc. Natl Acad. Sci. USA* **99**, 2918–2923 (2002).
- Lin, H. & Spradling, A. C. Germline stem cell division and egg chamber development in transplanted *Drosophila* germaria. *Dev. Biol.* **159**, 140–152 (1993).
- Clarkson, M. & Saint, R. A. His2AvDGFP fusion gene complements a lethal His2AvD mutant allele and provides an *in vivo* marker for *Drosophila* chromosome behaviour. *DNA Cell Biol.* **18**, 457–462 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Deshpande, T. Schupbach and A. Nouri for comments on the manuscript; A. Spradling, D. D. Barbosa, P. Ferree and W. Sullivan for stocks and reagents; B. Burdine for help with injection experiments; J. Goodhouse for microscopy assistance; and A. Basile, and the Wieschaus and Schupbach laboratory members for support during the realization of this work.

Author Contributions H.M.F. planned the project. H.M.F. and E.W. designed experiments. H.M.F. and J.M.L. performed experiments. D.N.R. wrote the image analysis software and statistical analysis. H.F. and E.W. contributed reagents and materials. H.M.F. wrote the paper. E.W., D.N.R. and J.M.L. edited the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H.M.F. (hfraydman@princeton.edu).

S-Nitrosylated protein-disulphide isomerase links protein misfolding to neurodegeneration

Takashi Uehara^{1,4}, Tomohiro Nakamura¹, Dongdong Yao¹, Zhong-Qing Shi¹, Zezong Gu¹, Yuliang Ma², Eliezer Masliah³, Yasuyuki Nomura⁴ & Stuart A. Lipton^{1,3}

Stress proteins located in the cytosol or endoplasmic reticulum (ER) maintain cell homeostasis and afford tolerance to severe insults^{1–3}. In neurodegenerative diseases, several chaperones ameliorate the accumulation of misfolded proteins triggered by oxidative or nitrosative stress, or of mutated gene products^{4,5}. Although severe ER stress can induce apoptosis^{2,6}, the ER withstands relatively mild insults through the expression of stress proteins or chaperones such as glucose-regulated protein (GRP) and protein-disulphide isomerase (PDI), which assist in the maturation and transport of unfolded secretory proteins. PDI catalyses thiol–disulphide exchange, thus facilitating disulphide bond formation and rearrangement reactions^{7–10}. PDI has two domains that function as independent active sites with homology to the small, redox-active protein thioredoxin^{7,8}. During neurodegenerative disorders and cerebral ischaemia, the accumulation of immature and denatured proteins results in ER dysfunction¹¹, but the upregulation of PDI represents an adaptive response to protect neuronal cells^{12–14}. Here we show, in brains manifesting sporadic Parkinson's or Alzheimer's disease, that PDI is S-nitrosylated, a reaction transferring a nitric oxide (NO) group to a critical cysteine thiol to affect protein function^{15–18}. NO-induced S-nitrosylation of PDI inhibits its enzymatic activity, leads to the accumulation of polyubiquitinated proteins, and activates the unfolded protein response. S-Nitrosylation also abrogates PDI-mediated attenuation of neuronal cell death triggered by ER stress, misfolded proteins or proteasome inhibition. Thus, PDI prevents neurotoxicity associated with ER stress and protein misfolding, but NO blocks this protective effect in neurodegenerative disorders through the S-nitrosylation of PDI.

Initially, we gathered two independent lines of chemical evidence to show that PDI was S-nitrosylated *in vitro* and *in vivo* to form an S-nitrosylated protein (SNO-P). First, in a specific fluorescence assay for SNO-P, we demonstrated the reaction of recombinant PDI with the physiological NO donor S-nitrosocysteine (SNOC; Fig. 1a). This assay detects the formation of SNO-P by the conversion of 2,3-diaminonaphthalene (DAN) to the fluorescent compound 2,3-naphthyltriazole (NAT)^{17,19,20}. SNOC-treated PDI resulted in significant SNO-P formation in a concentration-dependent manner.

Mammalian PDI has six cysteine residues in all, with four of them representing two thioredoxin-like domains (one near the amino terminus and the other near the carboxy terminus) that contain the Cys-Gly-His-Cys sequence at the active site. To determine the target site(s) of S-nitrosylation, we performed the DAN assay on immunoprecipitates from HEK-293T cell lysates transfected with wild-type or mutant PDI (cysteine-to-serine mutations in both of the active-site sequences). The fluorescence intensity in this assay of the N-terminal (C36S, C39S) and C-terminal (C383S, C386S) mutants

was decreased by about 50% compared with the wild type. The double mutant (N-terminal and C-terminal) was completely devoid of fluorescence, indicating that both thioredoxin-like domains are targets of S-nitrosylation (Fig. 1b). Next we showed that PDI in 293T cells was S-nitrosylated by the biotin-switch method. In this assay, a SNO-P is identified on western blots after replacing SNO with a more stable biotin group by chemical reduction with ascorbate, as described previously^{15,20,21}. SNOC markedly enhanced the level of S-nitrosylated PDI (SNO-PDI) in cell lysates or intact cells (Fig. 1c), whereas under the same conditions the NO[•] donor diethylamine-NO (DEA-NO) or hydrogen peroxide did not (Supplementary Fig. 1a). Additionally, in HEK-293 cells stably expressing neuronal NO synthase (nNOS), endogenous PDI was nitrosylated by endogenous NO, and this reaction was inhibited by a NOS inhibitor (Fig. 1d, e). Furthermore, by the biotin-switch assay we identified the cysteine residues that were S-nitrosylated. We found that endogenous nNOS activity led to the S-nitrosylation of cysteine residues in both thioredoxin-like domains of PDI (Fig. 1f and Supplementary Fig. 1b). Moreover, with mass spectrometry we found that one of the cysteine residues in the C-terminal thioredoxin-like domain was possibly further oxidized to sulphinic acid (–SO₂H) after exposure to NO (Supplementary Fig. 2). These data are consistent with our previous observation that reversible S-nitrosylation may facilitate further oxidation of the same cysteine thiol¹⁹.

Next, we sought to determine whether SNO-PDI was produced in neurodegenerative disorders associated with high levels of nitrosative stress and protein misfolding, such as Parkinson's disease (PD). First, we incubated dopaminergic SH-SY5Y cells with the mitochondrial complex I inhibitor rotenone, which is known to induce a parkinsonian phenotype, at least in part, in a NO-dependent fashion²². Exposure to rotenone led to the generation of SNO-PDI in these cells (Fig. 1g). To extend this finding to humans, we examined PD brains obtained shortly after death. We found evidence for SNO-PDI formation in each of four PD brains but not in controls obtained from patients who had died of disorders that were not of central nervous system origin (Fig. 1h and Supplementary Table 1). Additionally, brains from another major neurodegenerative disorder associated with protein aggregation and nitrosative stress, Alzheimer's disease (AD), also showed evidence of SNO-PDI (Fig. 1i, Supplementary Fig. 3 and Supplementary Table 1), consistent with the notion that this finding could represent a common denominator linking free-radical stress and protein misfolding.

To determine whether S-nitrosylation affects PDI function, we monitored PDI chaperone and isomerase activities. Chaperone activity was assessed in a standard assay by measuring the degree of aggregation of rhodanese induced by guanidinium, as previously described⁹. Rhodanese aggregation occurred in a time-dependent manner, and

¹Center for Neuroscience and Aging, and ²Proteomic Facility, Burnham Institute for Medical Research, 10901 North Torrey Pines Road, La Jolla, California 92037, USA.

³Department of Neurosciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92039, USA. ⁴Department of Pharmacology, Graduate School of Pharmaceutical Sciences, Hokkaido University, Sapporo, 060-0812, Japan.

incubation with recombinant wild-type PDI, but not dominant-negative PDI (produced by the N-terminal and C-terminal mutant)¹³, suppressed this aggregation by about 80%. S-Nitrosylation of PDI also significantly inhibited this chaperone activity (Fig. 2a). Next we measured isomerase activity in a standard assay that uses as a substrate an inactive form of RNase A containing scrambled disulphide bonds¹⁰. PDI catalyses the renaturation (refolding) of this inactive RNase A. Recovery of RNase A activity by wild-type PDI was attenuated about 50% by S-nitrosylation (Fig. 2b). Thus, S-nitrosylation inhibited the functional activities of PDI. In addition, direct oxidation of PDI by hydrogen peroxide could also decrease its activity (Supplementary Fig. 4), indicating

that the sulphinated PDI derivative observed by mass spectrometry after exposure to NO (Supplementary Fig. 2) might also be pathophysiologically relevant to the inhibition of PDI activity.

From these results we reasoned that PDI might function in attenuating protein misfolding and consequent aggregation in

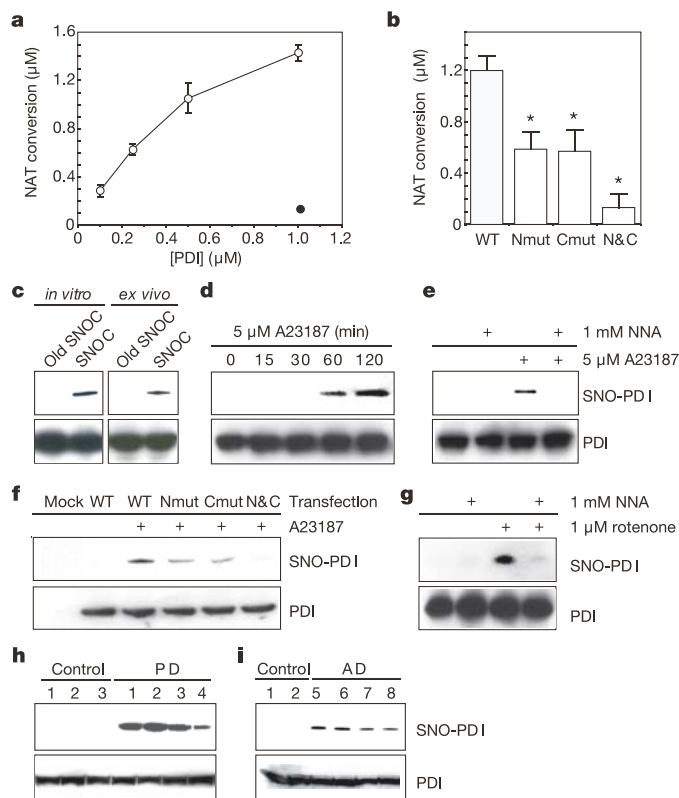


Figure 1 | S-Nitrosylation of PDI *in vitro* and *in vivo*. **a**, Recombinant PDI protein was incubated with SNOC (100 μ M) for 30 min at room temperature 21 °C. SNO-PDI thus generated was assessed by DAN assay ($n = 4$ experiments). Open symbols, with SNOC; filled symbol, without SNOC. Values are mean $\pm$ s.e.m. **b**, Cell lysates transduced with wild-type (WT) or mutant (N-terminal (Nmut), C-terminal (Cmut), or N-terminal and C-terminal (N&C)) PDI were immunoprecipitated with anti-PDI antibody. Immunoprecipitates were then incubated in the presence or absence of SNOC and subjected to DAN assay. Values are mean $\pm$ s.e.m., $n = 5$; asterisks, $P < 0.01$ by analysis of variance. **c**, Top: cell lysates from human 293T cells were incubated with SNOC at room temperature to assay for SNO-PDI. Control samples were subjected to decayed (old) SNOC. SNO-PDI was detected by biotin-switch assay 30 min after SNOC exposure. Bottom: total PDI in cell lysates by western analysis. **d–f**, Top panels: HEK-293 cells stably expressing nNOS were assayed for endogenous SNO-PDI. nNOS was activated by Ca^{2+} ionophore A23187 (5 μ M) in the presence or absence of NOS inhibitor (N-nitro-L-arginine; NNA). Bottom panels: Total PDI. Activation of nNOS increased endogenous SNO-PDI (**d**). NNA prevented this increase (**e**). Mutation of critical cysteine thiol groups of PDI also prevented its S-nitrosylation (**f**). **g**, SNO-PDI increased in cells exposed to rotenone in a NOS-dependent manner. Top: lysates from SH-SY5Y cells exposed to rotenone for 6 h in the presence or absence of NNA. Bottom: Total PDI. **h, i**, Brain tissues from controls, from PD patients with diffuse Lewy body disease (**h**), or AD patients (**i**) were subjected to biotin-switch assay to detect *in vivo* S-nitrosylation (see additional results in Supplementary Fig. 3 and Supplementary Table 1).

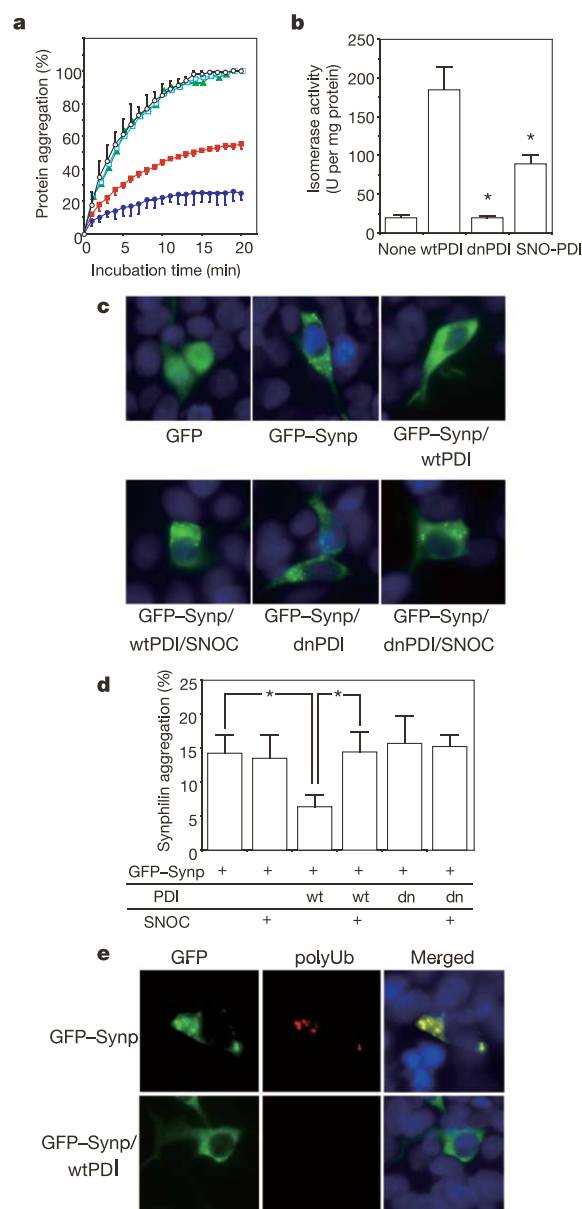


Figure 2 | S-Nitrosylation of PDI regulates its enzymatic activity. **a**, Effect of S-nitrosylation on PDI chaperone activity. Unfolded rhodanese was diluted in buffer containing wild-type PDI (blue), SNO-PDI (red), SNOC (green), dominant-negative PDI (N-terminal and C-terminal mutant; cyan) or buffer alone (black). Rhodanese aggregation was monitored ($n = 8$ experiments). Values are mean $\pm$ s.e.m. **b**, S-Nitrosylation of PDI attenuates its isomerase activity. Scrambled RNase A from bovine pancreas was incubated with wild-type (wt) PDI, dominant-negative (dn) PDI or SNO-PDI. Asterisks, $P < 0.01$ for $n = 5$ experiments. Values are mean $\pm$ s.e.m. **c**, PDI inhibits the aggregation of synphilin-1. SH-SY5Y cells were transfected with green fluorescent protein (GFP)–synphilin-1 (GFP–Synp) and wild-type or dominant-negative PDI. Inclusion body formation was monitored by deconvolution microscopy 24 h after exposure to SNOC or control solution. Images were deconvolved with SlideBook software (Intelligent Imaging Innovations, Inc.). **d**, Percentage of cells with GFP–Synp inclusions. Values are mean $\pm$ s.e.m. for $n = 2,500$ transfectants counted in five experiments; asterisks, $P < 0.05$. **e**, Co-localization of synphilin-1 and polyubiquitin (polyUb) by immunofluorescence ($n = 3$).

neurodegenerative diseases, and that the formation of SNO-PDI might inhibit neuroprotective activity. Moreover, previous work had suggested that ER stress can directly or indirectly influence the aggregation of both ER and cytosolic proteins^{4,5,23}. We therefore attempted to show an inhibitory effect of PDI on the aggregation of synphilin-1, as observed in Lewy body inclusions in the brains of PD patients. Initially, we tested whether PDI could prevent the ubiquitinated, Lewy-body-like inclusions that are formed in the cytosol after synphilin-1 overexpression in cultured SH-SY5Y cells (Fig. 2c). When wild-type PDI was co-expressed with synphilin-1, discrete inclusions were greatly decreased, and instead ubiquitin-negative synphilin-1 was localized diffusely in the cytosol (Fig. 2c–e). As a control, immunofluorescent staining of transfected SH-SY5Y cells revealed that overexpression of PDI did not alter its predominant intracellular distribution in the ER (Supplementary Fig. 5). NO attenuated the protective effect of PDI on synphilin-1 inclusions (Fig. 2d). These findings suggest that PDI is involved in protein folding linked to PD in a NO-sensitive manner.

Excitotoxic damage is also thought to have a function in neurodegenerative disorders such as PD by triggering the production of free radicals, including NO, in part through the excessive stimulation of *N*-methyl-D-aspartate (NMDA)-type glutamate receptors; mild exposure to NMDA results in neuronal apoptosis^{24–27}. Here we found that exposure of cerebrocortical neurons to NMDA induced

SNO-PDI in a NOS-sensitive fashion (Fig. 3a). We surmised that abrogation of PDI chaperone and enzymatic activity by *S*-nitrosylation might contribute to the accumulation of unfolded and consequently polyubiquitinated proteins marked for degradation by the proteasome. We therefore next examined whether the accumulation of polyubiquitinated proteins occurred in response to NMDA by using a polyubiquitin-specific antibody. Within 12 h of exposure to NMDA, we detected polyubiquitin immunoreactivity in neurons, but the cells remained viable at this time point (Fig. 3b, c). By 24 h, many of the polyubiquitinated neurons had undergone apoptosis. Overexpression of wild-type PDI decreased the number of apoptotic, polyubiquitinated cells, indicating that PDI has a function in preventing the accumulation of unfolded, polyubiquitinated proteins in response to NMDA insult, and subsequent neuronal cell death. Next we evaluated the involvement of the unfolded protein response (UPR) signalling pathway that is activated by the accumulation of misfolded proteins or ER dysfunction. Representing this pathway we detected *CHOP* mRNA induction and *XBP-1* mRNA processing by activated IRE1- α after exposure of cortical cultures to

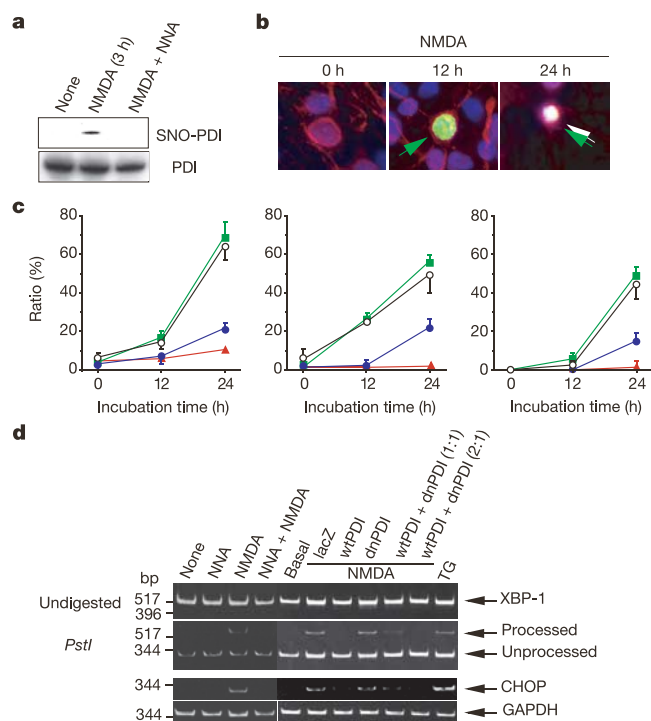


Figure 3 | NMDA stimulates the accumulation of polyubiquitinated proteins and UPR pathway. **a**, SNO-PDI was detected in a NOS-sensitive manner in primary cortical cultures exposed to NMDA. **b**, Primary cortical cells infected with adenoviral Ad-LacZ or Ad-wtPDI were exposed to NMDA and immunostained for polyubiquitinated protein (green) and neuron-specific MAP2 (red). Hoechst-stained DNA (blue) was used to assess condensed, apoptotic nuclei (white in merged image). **c**, Quantification of apoptotic and/or polyubiquitinated neurons similar to those shown in **b**. Left, apoptotic cells; middle, polyubiquitinated cells; right, apoptotic and polyubiquitinated cells. Values are ratio ($\times 100\%$) of affected to total neurons, expressed as mean $\pm$ s.e.m. for $n = 7,000$ neurons counted in six experiments; asterisks, $P < 0.01$. Open circles, Ad-LacZ; blue circles, Ad-wtPDI; green squares, Ad-dnPDI; red triangles, Ad-wtPDI plus *N*-nitro-L-arginine. **d**, NMDA-stimulated processing of *XBP-1* mRNA and induction of *CHOP*. Processing of endogenous *XBP-1* mRNA was evaluated by *Pst*I endonuclease digestion of *XBP-1* cDNA. TG, thapsigargin; bp, base pairs.

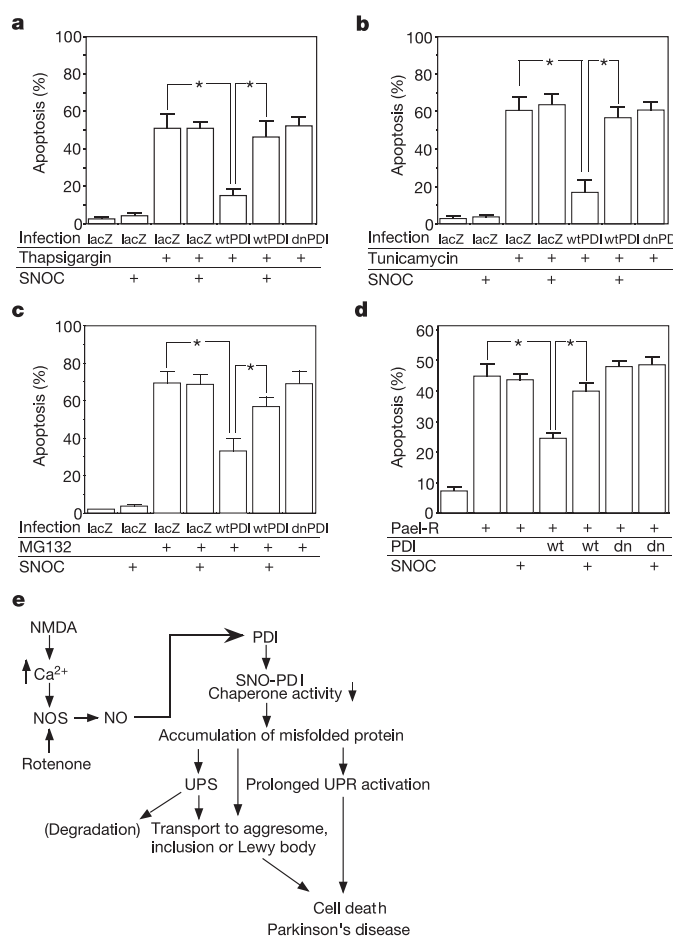


Figure 4 | Neuroprotection by PDI against ER stress, proteasome inhibition, or Pael receptor expression. **a–d**, SH-SY5Y cells were transfected for 24 h with expression vectors for control lacZ (**a–c**), PDI constructs (**a–d**), or Pael receptor (**d**). Cultures were then incubated for 15 h in the presence or absence of 100 μ M SNOC with 5 μ M thapsigargin (**a**), 10 μ M tunicamycin (**b**) or 0.1 μ M proteasome inhibitor MG132 (**c**). Exposure to SNOC abolished the protective effect of PDI on cell death induced by ER stress, proteasome inhibitor or Pael receptor. dn, dominant-negative; wt, wild type. For each panel, values are mean $\pm$ s.e.m. for $n = 3,500$ cells counted in five experiments; asterisks, $P < 0.01$. **e**, Possible mechanism of SNO-PDI contributing to the accumulation of aberrant proteins and to cell death in human neurodegenerative disorders. UPS, ubiquitin-proteasome system.

NMDA^{1,2}. This UPR was attenuated by overexpression of wild-type PDI but not by dominant-negative PDI (Fig. 3d). Additionally, the NOS inhibitor *N*-nitro-L-arginine blocked NMDA-induced apoptosis and the UPR, indicating that a pathophysiologically relevant amount of NO was produced under these conditions (Fig. 3c, d and Supplementary Fig. 6). Taken together, these findings indicate that NMDA activates a NO-mediated UPR through ER dysfunction, but this dysfunction can be mitigated by PDI activity.

For further clarification of the relationship between the protective function of PDI and its *S*-nitrosylation, we investigated the effect of PDI on neuronal death after ER stress or proteasome inhibition (resulting in the accumulation of polyubiquitinated proteins that cannot be degraded by the proteasome). For this purpose we used SH-SY5Y cells because, unlike cortical neurons, they were resistant to direct NO-induced damage under conditions of SNO-PDI formation (Fig. 1), allowing us to tease apart the effect of NO on cell death and PDI *S*-nitrosylation. We found that cell death precipitated by thapsigargin and tunicamycin (to induce ER stress) or MG132 (to inhibit the proteasome) was largely abrogated by wild-type PDI; however, this protective effect was reversed by exposure to SNO (Fig. 4a–c and Supplementary Fig. 6). Similarly, wild-type PDI ameliorated cell death triggered by overexpression of the Pael receptor, a protein that abnormally accumulates in Parkinson's disease and serves as a potent inducer of the UPR and substrate of the E3 ubiquitin ligase parkin^{6,28}; exposure to SNO also reversed this protective effect (Fig. 4d). These results are consistent with the notion that NO impairs the protective role of PDI through *S*-nitrosylation. From these findings we conclude that cell death in response to proteasome inhibition or ER stress, which contributes to ER dysfunction, UPR activation and protein misfolding, can be attenuated by PDI.

These results show that SNO-PDI forms in brains of patients with PD and AD, neurodegenerative disorders that are characterized by abnormal protein accumulations. Cell models of neurodegeneration produced by exposure to the pesticide rotenone, NO or NMDA also result in the formation of SNO-PDI. *S*-Nitrosylation of PDI inhibits its activity, allows the accumulation of polyubiquitinated proteins and contributes to neuronal cell death (Fig. 4e). To determine whether the level of SNO-PDI in neurodegenerative human brain is of pathophysiological significance, we calculated the ratio of SNO-PDI (by biotin-switch assay) to total PDI (from western blotting) and found that this ratio was similar to that encountered in our neuronal cell models manifesting polyubiquitinated proteins and cell death (Supplementary Fig. 3). This finding indicates that pathophysiologically relevant amounts of SNO-PDI are present in human brains with PD and AD.

In the absence of nitrosylation, wild-type PDI attenuates abnormal protein accumulation and ubiquitination, including the parkinsonian-related protein synphilin-1, and affords neuroprotection. Previous reports have shown the development of ER stress and UPR activation in cellular models of PD²⁹. In addition, there is increasing evidence that the accumulation of aggregated or misfolded proteins links cellular stress to the pathogenesis of PD^{6,20,30}. Other reports have shown that NO can be involved in neurodegeneration by a variety of mechanisms^{15,19,20,24–27,30}. Our data demonstrate a previously unrecognized relationship between NO and protein misfolding in degenerative disorders, showing that PDI can be a target of NO after mitochondrial insult in cellular models of PD and in human neurodegenerative diseases. Nitrosative stress resulting in PDI dysfunction therefore provides a mechanistic link between deficits in molecular chaperones, accumulation of misfolded proteins, and neuronal demise in neurodegenerative disorders. The elucidation of this SNO-PDI-mediated pathway that contributes to neuronal injury and apoptosis might permit the development of new therapeutic approaches for neurodegenerative diseases and other disorders associated with abnormal protein accumulation and nitrosative stress.

METHODS

Fluorimetric detection of *S*-nitrosothiols. We measured *S*-nitrosothiols by the conversion of DAN to fluorescent NAT, as described^{17,19}. NAT was quantified with a FluoroMax-2 spectrofluorometer and DataMax software (Instruments S.A.). Serial NAT dilutions were used to construct a standard curve.

Cell injury/death assays. Cerebrocortical neurons or SH-SY5Y cells were transduced with a wild-type or mutated PDI gene or with a Pael receptor expression construct and incubated for 24 h (Supplementary Fig. 7). The cells were then treated for 15 h with 5 μ M thapsigargin, 10 μ g ml^{−1} tunicamycin, 0.1 μ M MG-132 or diluent in the presence or absence of 100 μ M SNO. Neurotoxicity of the Pael receptor was analysed as described⁶. Cortical neurons exposed to NMDA were incubated as described previously²⁴. Hoechst staining was used to assess morphological changes of apoptotic nuclei. MAP2 staining was used to assess injury or retraction of neuronal processes. Additional details are described in Supplementary Information.

Expression and purification of recombinant PDI proteins, isolation of PDI complementary DNA and construction of adenoviral vectors, detection of *S*-nitrosylated proteins with the biotin-switch assay, liquid chromatography–mass spectrometry of PDI, cell culture and transduction by means of adenovirus, PDI enzymatic activity assays, detection of aggregated synphilin-1, immunocytochemistry, *XBP-1* mRNA splicing and *CHOP* mRNA induction, human subjects, and statistics are all described in the Supplementary Information.

Received 27 December 2005; accepted 4 April 2006.

- Elgaard, L., Molinari, M. & Helenius, A. Setting the standards: quality control in the secretory pathway. *Science* **286**, 1882–1888 (1999).
- Kaufman, R. J. Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls. *Genes Dev.* **13**, 1211–1233 (1999).
- Patil, C. & Walter, P. Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. *Curr. Opin. Cell Biol.* **13**, 349–355 (2001).
- Rao, R. V. & Bredesen, D. E. Misfolded proteins, endoplasmic reticulum stress and neurodegeneration. *Curr. Opin. Cell Biol.* **16**, 653–662 (2004).
- Haynes, C. M., Titus, E. A. & Cooper, A. A. Degradation of misfolded proteins prevents ER-derived oxidative stress and cell death. *Mol. Cell* **15**, 767–776 (2004).
- Imai, Y. *et al.* An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin. *Cell* **105**, 891–902 (2001).
- Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A. & Rutter, W. J. Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin. *Nature* **317**, 267–270 (1985).
- Vuori, K., Myllylä, R., Pihlajaniemi, T. & Kivirikko, K. I. Expression and site-directed mutagenesis of human protein disulfide isomerase in *Escherichia coli*. This multifunctional polypeptide has two independently acting catalytic sites for the isomerase activity. *J. Biol. Chem.* **267**, 7211–7214 (1992).
- Song, J. L. & Wang, C. C. Chaperone-like activity of protein disulfide-isomerase in the refolding of rhodanese. *Eur. J. Biochem.* **231**, 312–316 (1995).
- Gilbert, H. F. Protein disulfide isomerase. *Methods Enzymol.* **290**, 26–50 (1998).
- Hu, B. R., Martone, M. E., Jones, Y. Z. & Liu, C. L. Protein aggregation after transient cerebral ischemia. *J. Neurosci.* **20**, 3191–3199 (2000).
- Ko, H. S., Uehara, T. & Nomura, Y. Role of ubiquitin associated with protein-disulfide isomerase in the endoplasmic reticulum in stress-induced apoptotic cell death. *J. Biol. Chem.* **277**, 35386–35392 (2002).
- Tanaka, S., Uehara, T. & Nomura, Y. Up-regulation of protein-disulfide isomerase in response to hypoxia/brain ischemia and its protective effect against apoptotic cell death. *J. Biol. Chem.* **275**, 10388–10393 (2000).
- Conn, K. J. *et al.* Identification of the protein disulfide isomerase family member PDIP in experimental Parkinson's disease and Lewy body pathology. *Brain Res.* **1022**, 164–172 (2004).
- Lipton, S. A. *et al.* A redox-based mechanism for the neuroprotective and neurodestructive effects of nitric oxide and related nitroso-compounds. *Nature* **364**, 626–632 (1993).
- Stamler, J. S. Redox signaling: nitrosylation and related target interactions of nitric oxide. *Cell* **78**, 931–936 (1994).
- Haendeler, J. *et al.* Redox regulatory and anti-apoptotic functions of thioredoxin depend on *S*-nitrosylation at cysteine 69. *Nature Cell Biol.* **4**, 743–749 (2002).
- Sliskovic, I., Raturi, A. & Mutus, B. Characterization of the *S*-denitrosation activity of protein disulfide isomerase. *J. Biol. Chem.* **280**, 8733–8741 (2005).
- Gu, Z. *et al.* *S*-Nitrosylation of matrix metalloproteinases: signaling pathway to neuronal cell death. *Science* **297**, 1186–1190 (2002).
- Yao, D. *et al.* Nitrosative stress linked to sporadic Parkinson's disease: *S*-nitrosylation of parkin regulates its E3 ubiquitin ligase activity. *Proc. Natl Acad. Sci. USA* **101**, 10810–10814 (2004).
- Jaffrey, S. R., Erdjument-Bromage, H., Ferris, C. D., Tempst, P. & Snyder, S. H.

- Protein S-nitrosylation: a physiological signal for neuronal nitric oxide. *Nature Cell Biol.* **3**, 193–197 (2001).
22. Betarbet, R. *et al.* Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nature Neurosci.* **3**, 1301–1306 (2000).
 23. Nishikawa, A. *et al.* Novel function of PS2V: change in conformation of tau proteins. *Biochem. Biophys. Res. Commun.* **318**, 435–438 (2004).
 24. Bonfoco, E., Krainc, D., Ankarcrona, M., Nicotera, P. & Lipton, S. A. Apoptosis and necrosis: two distinct events induced, respectively, by mild and intense insults with N-methyl-D-aspartate or nitric oxide/superoxide in cortical cell cultures. *Proc. Natl Acad. Sci. USA* **92**, 7162–7166 (1995).
 25. Dawson, V. L., Dawson, T. M., London, E. D., Bredt, D. S. & Snyder, S. H. Nitric oxide mediates glutamate neurotoxicity in primary cortical cultures. *Proc. Natl Acad. Sci. USA* **88**, 6368–6371 (1991).
 26. Lipton, S. A. & Rosenberg, P. A. Excitatory amino acids as a final common pathway for neurologic disorders. *N. Engl. J. Med.* **330**, 613–622 (1994).
 27. Hara, M. R. *et al.* S-nitrosylated GAPDH initiates apoptotic cell death by nuclear translocation following Siah1 binding. *Nature Cell Biol.* **7**, 665–674 (2005).
 28. Murakami, T. *et al.* Pael-R is accumulated in Lewy bodies of Parkinson's disease. *Ann. Neurol.* **55**, 439–442 (2004).
 29. Ryu, E. J. *et al.* Endoplasmic reticulum stress and the unfolded protein response in cellular models of Parkinson's disease. *J. Neurosci.* **22**, 10690–10698 (2002).
 30. Chung, K. K. *et al.* S-Nitrosylation of parkin regulates ubiquitination and compromises parkin's protective function. *Science* **304**, 1328–1331 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank X. Fang for preparation of cerebrocortical cultures, T. William for technical assistance with the analysis of mass spectra, and R. Takahashi for the Pael receptor construct. T.U. was supported in part by the Mitsubishi Pharma Research Foundation and a Grant-in-Aid from the Ministry of Education, Culture, Sports and Technology of Japan. S.A.L. was supported in part by grants from the NIH, the American Parkinson's Disease Association, San Diego Chapter, and an Ellison Senior Scholars Award in Aging.

Author Contributions T.U. and T.N. performed most of the experiments, contributing equally to the work, and helped to write the manuscript. D.Y., Z.Q.S. and Z.G. provided the biochemical data, and also contributed equally to the work. Y.M. analyzed the mass spectrometry data. E.M. provided the human subjects, and Y.N. provided constructs and advice. S.A.L., the senior author, designed the project, helped to analyse the data, wrote the manuscript and provided the financial support.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare that they have no competing financial interests. Correspondence and requests for materials should be addressed to S.A.L. (slipton@burnham.org).

LETTERS

PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention

Jiawang Zhang¹, Justin C. Grindley¹, Tong Yin¹, Sachintha Jayasinghe¹, Xi C. He¹, Jason T. Ross¹, Jeffrey S. Haug¹, Dawn Rupp¹, Kimberly S. Porter-Westpfahl¹, Leanne M. Wiedemann^{1,3}, Hong Wu² & Linheng Li^{1,3}

Haematopoietic stem cells (HSCs) must achieve a balance between quiescence and activation that fulfils immediate demands for haematopoiesis without compromising long-term stem cell maintenance, yet little is known about the molecular events governing this balance^{1–3}. Phosphatase and tensin homologue (PTEN) functions as a negative regulator of the phosphatidylinositol-3-OH kinase (PI(3)K)–Akt pathway, which has crucial roles in cell proliferation, survival, differentiation and migration^{4,5}. Here we show that inactivation of PTEN in bone marrow HSCs causes their short-term expansion, but long-term decline, primarily owing to an enhanced level of HSC activation. PTEN-deficient HSCs engraft normally in recipient mice, but have an impaired ability to sustain haematopoietic reconstitution, reflecting the dysregulation of their cell cycle and decreased retention in the bone marrow niche. Mice with PTEN-mutant bone marrow also have an increased representation of myeloid and T-lymphoid lineages and develop myeloproliferative disorder (MPD)⁶. Notably, the cell populations that expand in PTEN mutants match those that become dominant in the acute myeloid/lymphoid leukaemia that develops in the later stages of MPD. Thus, PTEN has essential roles in restricting the activation of HSCs, in lineage fate determination, and in the prevention of leukaemogenesis.

The *Pten* tumour suppressor gene is commonly mutated in malignancy, including leukaemias that feature dysregulated haematopoiesis^{4,5}. To ablate PTEN function in adult HSCs, we crossed a polyinosine-polycytidine (pIpC)-inducible *Mx-1-Cre* recombinase mouse line^{7,8} with a conditional PTEN-mutant (*Pten^{fl/fl}*) mouse line⁹, and induced Cre expression with pIpC on postnatal day 21, 23 and 25 (Supplementary Fig. 1a). Times measured from the final pIpC injection are expressed as days post-induction (DPI). Our study focuses on *Mx-1-Cre⁺;Pten^{fl/fl}* animals, hereafter referred to as 'PTEN mutants'. Control animals were pIpC-treated *Mx-1-Cre⁻;Pten^{fl/fl}* mice, which are phenotypically equivalent to animals with a wild-type *Pten* allele (Supplementary Fig. 1b).

At 30 DPI, we found that PTEN-mutant mice had fewer HSCs than controls, as measured by the absolute number of Lin[−]Sca-1⁺c-Kit⁺ cells (referred to as 'LSK' cells) in the bone marrow using flow cytometry. LSK cells are a heterogeneous population that includes long-term HSCs (LT-HSCs; LSK Flk-2[−]) and short-term HSCs (ST-HSCs; LSK Flk-2⁺), which are capable of sustaining haematopoiesis for many months, or for a few weeks, respectively¹⁰. The decrease in PTEN-mutant HSCs stemmed from a decline in the LT-HSC population, whereas the ST-HSC population was unchanged (Fig. 1a). Thus, PTEN is required for LT-HSC maintenance.

PTEN controls cell cycle entry and progression through the inhibition of PI(3)K–Akt activity^{4,5}, so we examined the cell cycle profiles of the HSCs—as determined by RNA versus DNA content in

LSK cells. PTEN mutants showed a 2–3-fold decrease in the proportion of HSCs in the G0 phase and a concomitant increase in the proportion of HSCs in the S + G2M phases (Fig. 1b). Consistent with this result, staining for the proliferation antigen Ki67 revealed that the proportion of cycling HSCs is aberrantly high in PTEN mutants (Supplementary Fig. 1c).

Next, we examined the relationship between PTEN expression and cell cycle status. PTEN is phosphorylated at its carboxy terminus by casein kinase II, producing a phospho-PTEN (p-PTEN) form that is unable to be recruited to the plasma membrane to antagonize PI(3)K–Akt activity¹¹. *In vivo* p-PTEN is indicative of cells in which Akt is in an active state¹², so we distinguished between p-PTEN and

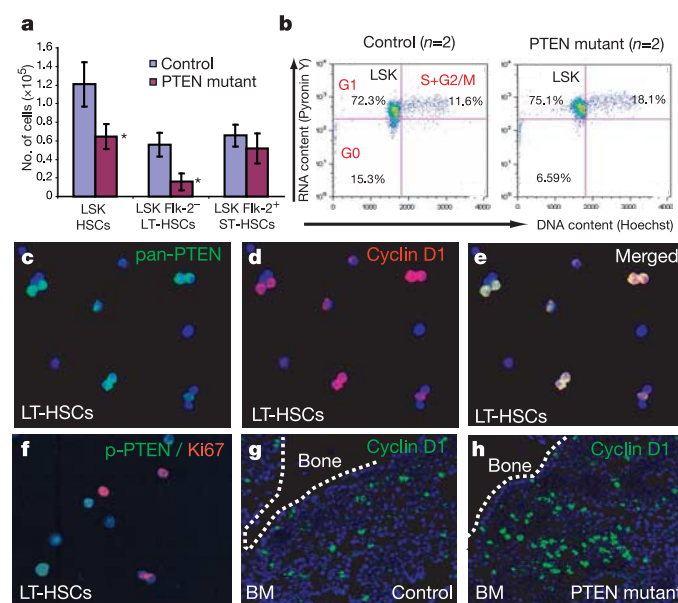


Figure 1 | PTEN governs the number and activation state of HSCs and PTEN expression is associated with cyclin D1. **a**, Absolute number of HSCs (LSK), LT-HSCs (LSK Flk-2[−]) and ST-HSCs (LSK Flk-2⁺) in control and PTEN-mutant bone marrow at 30 DPI. *, *P* < 0.05. **b**, Comparison of LSK cells in G0, G1 or S+G2/M phases between control animals and PTEN mutants. **c–e**, Sorted LSK Flk-2[−] preparations co-stained with antibodies that recognize pan-PTEN (green; **c**, **e**) and cyclin D1 (red; **d**, **e**), and counterstained with 4,6-diamidino-2-phenylindole (DAPI; blue). **e**, Co-staining with antibodies that recognize p-PTEN (green) and Ki67 (red), indicating that most p-PTEN⁺ cells are Ki67⁺. **g**, **h**, Bone marrow (BM) sections from control (**g**) and PTEN-mutant (**h**) mice stained with anti-cyclin D1 antibody (green). Error bars in **a** represent s.d.

¹Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA. ²Department of Molecular and Medical Pharmacology, UCLA School of Medicine, Los Angeles, California 90095, USA. ³Department of Pathology and Laboratory Medicine, Kansas University Medical Center, Kansas City, Kansas 66160, USA.

the non-phosphorylated form (non-p-PTEN) by staining with one antibody that recognizes both forms of PTEN (pan-PTEN), and another antibody that recognizes only p-PTEN. Performing immunofluorescence in sorted LT-HSCs, we observed a strong association between the pan-PTEN signal and the presence of cyclin D1. The pan-PTEN signal was found almost exclusively (98% of the time) in cells that were also cyclin D1 positive (cyclin D1⁺), and most (82%) cyclin D1⁺ cells were also pan-PTEN⁺ (Fig. 1c–e; Supplementary Fig. 1d). Cyclin D1, a known target of the PI(3)K–Akt pathway⁴, maintains cells at the G1 phase in preparation for the G1/S-phase transition, so these data indicate that PTEN is regulated through the cell cycle and is predominantly expressed in cells at G1. In contrast, only around half of the cyclin D1⁺ cells were found to be p-PTEN⁺ (data not shown), although p-PTEN was rarely present in cycling (Ki67⁺) HSCs (Fig. 1f). We also observed that PTEN deficiency markedly increased the number of cyclin D1⁺ cells in bone marrow sections (Fig. 1g, h), although this probably reflects the combined impact of the mutation on HSCs and on multiple lineages of their progeny.

HSCs are normally located in bone marrow niches^{3,7,13,14}. As the PTEN–Akt pathway can regulate cell migration⁴, we asked whether the lack of PTEN affects the retention of HSCs in the bone marrow. We used flow cytometry at 2, 6, 9, 12 and 30 DPI to examine HSC (LSK cell) numbers in the bone marrow, peripheral blood and spleen. In addition, at the 30-day time point we used quantitative colony-forming unit (CFU) assays (see Supplementary Methods) to functionally assess the numbers of either HSCs (by measuring CFU-spleen, or CFU-S) or multiple myeloid progenitors (measuring CFU in culture, or CFU-C). We found that PTEN deficiency led to a modest increase in the percentage of bone marrow HSCs at 2 DPI, but a decline from 6 DPI onwards (Fig. 2a). These changes were accompanied by dramatic increases in the HSC population in the peripheral blood and spleen (for example, a 10-fold increase on day 30 in the spleen; Fig. 2a, middle and lower panels; Supplementary Fig. 2a, b, d). Thus, PTEN mutation leads to mobilization of HSCs from the bone marrow into the peripheral blood and spleen. CFU-S assays confirmed this result. PTEN mutant-derived bone marrow contained fewer CFU-S than control bone marrow; PTEN-mutant spleen cells formed CFU-S, whereas those derived from control mice did not (Fig. 2b; Supplementary Fig. 2c). HSC mobilization can result in extramedullary (outside the bone marrow) haematopoiesis when the bone marrow is stressed. Indeed, we found that in PTEN-mutant mice progenitor cells were present ectopically in the peripheral blood and spleen. The number of peripheral blood- or spleen-derived CFU-C was substantially increased in the PTEN mutants compared with controls, whereas bone marrow-derived CFU-C remained unchanged (Fig. 2c; Supplementary Fig. 2e).

Exploring the mechanism of HSC mobilization¹⁵, we found that PTEN deficiency did not affect the ability of HSCs to adhere to fibronectin, collagen or laminin (Supplementary Fig. 3c), nor was there any significant alteration in migration speed (data not shown) or migrating cell number (Supplementary Fig. 3b) when PTEN-mutant cells were exposed to stromal cell-derived factor 1 (SDF-1; also known as CXCL12). Moreover, in PTEN mutants there was no change in the expression of either CXCR4 (an SDF-1 receptor) or $\alpha_{4\beta 1}$ integrin, both of which are important for HSC homing and migration^{15–17} (Supplementary Fig. 3a).

We then performed *in vivo* homing and lodging assays, in which fluorescently labelled bone marrow is transplanted into wild-type hosts and assessed for the ability to home to or lodge within the bone marrow (primarily a property of HSCs). We found that PTEN deficiency did not affect homing to the bone marrow of irradiated hosts (a treatment that destroys resident HSCs, resulting in vacant niches and elevated levels of SDF-1) (Fig. 2d). In contrast, when transplanted into non-irradiated hosts (that is, having intact bone marrow with few vacant niches and baseline SDF-1 levels), only half as many mutant as control cells lodged in the bone marrow at

6-hours post-transplantation, and this difference was greater still by 18-hours post-transplantation (Fig. 2e). Thus, PTEN HSCs are capable of reaching and residing in the bone marrow when conditions are conducive to homing, but they do not perform as well as normal HSCs when there are few vacant niches and competing wild-type host HSCs are present.

HSCs that mobilize to the spleen are more proliferative than HSCs

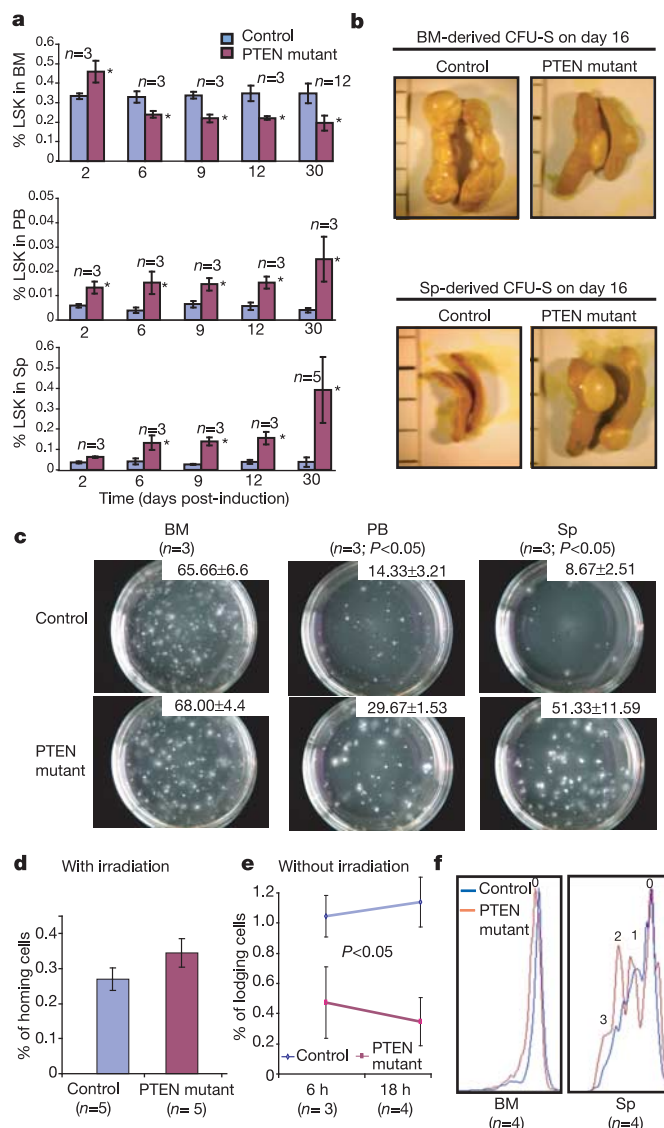


Figure 2 | Deletion of PTEN results in HSC expansion and mobilization and impaired bone marrow lodging. **a**, Flow cytometry analysis of LSK cells on the indicated days post-induction in control and PTEN-mutant bone marrow (BM; upper panel), peripheral blood (PB; middle panel) and spleen (Sp; lower panel). *, $P < 0.05$. **b**, Comparison of CFU-S on day 16 post-transplantation for control and PTEN-mutant bone marrow (upper panels) and spleen (lower panels). **c**, Comparison of *in vitro* CFU-C (a combination of CFU-E (erythroid), CFU-GM (granulocyte, macrophage) and CFU-GEMM (granulocyte, erythrocyte, monocyte, megakaryocyte)) in control and PTEN-mutant bone marrow (left panels), peripheral blood (middle panels) and Sp (right panels) counted on day 12. **d**, Analysis of bone marrow homing ability in control and PTEN-mutant bone marrow in an irradiated host mouse. **e**, **f**, Analysis of the bone marrow lodging ability of control and PTEN-mutant bone marrow in a non-irradiated host (e) and the corresponding proliferation profiles of lodged cells assessed by the dilution of their fluorescent label (f). Peaks in f represent cell populations that underwent the number of cell divisions (0, 1, 2 or 3) as shown. Error bars in a, d, e represent s.d.

retained in the bone marrow¹⁸. If the ability to lodge in the bone marrow is linked to cell cycle status, the reduced lodging of PTEN-mutant cells may reflect their increased likelihood of being in an activated state. Indeed, our lodging assays showed a strong association between proliferation rate and lodging location (Fig. 2f). Irrespective of PTEN status, cells that lodged in the bone marrow were predominantly quiescent (no cell division) up to the 18-hour time point. In contrast, of the cells that were lodged in the spleen at the same time point, 60% of the PTEN mutants versus 45% of the controls ($P = 0.0042$) had undergone one, two or three cell divisions (compare peaks of the blue and red curves, Fig. 2f).

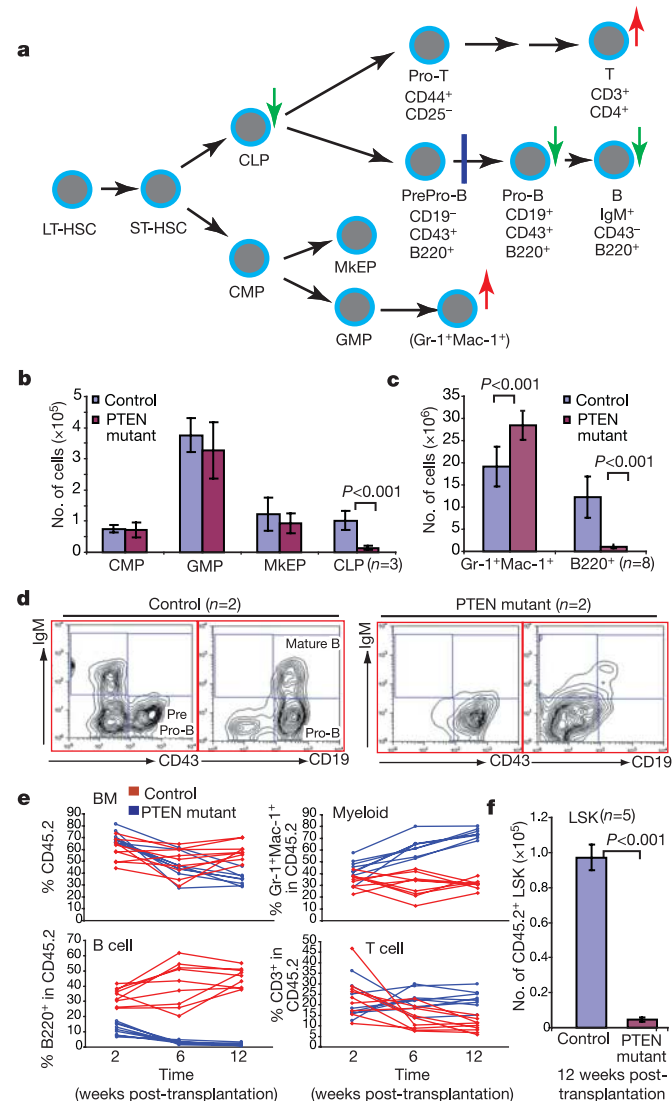


Figure 3 | Lineage analysis and competitive repopulation assay.

a, Haematopoietic lineage tree showing the impact of PTEN deficiency. Increased and decreased cell populations are marked by red and green arrows, respectively. The blue vertical bar indicates a point of developmental blockade. See the text for definitions of the cell populations. **b**, **c**, Absolute numbers of CLP, CMP, GMP and MKEP cells (**b**) and Gr-1⁺Mac-1⁺ myeloid and B220⁺ B-lineage populations (**c**) in control and PTEN-mutant bone marrow at 30 DPI. **d**, Analysis of B-lineage development, profiling sorted B220⁺ cell populations from controls and PTEN mutants. **e**, **f**, Competitive repopulation analysis of control and PTEN-mutant bone marrow. Recipient mice were transplanted with 2×10^5 rescue bone marrow cells together with 1×10^6 PTEN-mutant bone marrow or 2×10^5 control bone marrow cells. Analyses of the donor (CD45.2⁺) contribution to different lineages in the peripheral blood (**e**) and to HSCs in the bone marrow (**f**). Error bars in **b**, **c**, **f** represent s.d.

During haematopoietic cell development, HSCs progressively give rise to different progenitor cells, which, in turn, generate the various mature lineages^{10,19,20} (see also Fig. 3a). Thus, we determined the impact of PTEN deficiency on the haematopoietic lineage at 30 DPI, and found that CLP (common lymphoid progenitor) numbers were significantly lower in PTEN-mutant bone marrow compared with control bone marrow, whereas the CMP (common myeloid progenitor), GMP (granulocyte, monocyte progenitor) and MKEP (megakaryocyte, erythroid progenitor) populations remained unchanged (Fig. 3b). Although the number of Gr-1⁺Mac-1⁺ myeloid cells was increased, the number of B lymphocytes (B220⁺) was substantially decreased (Fig. 3c), the latter change involving both the loss of CLPs

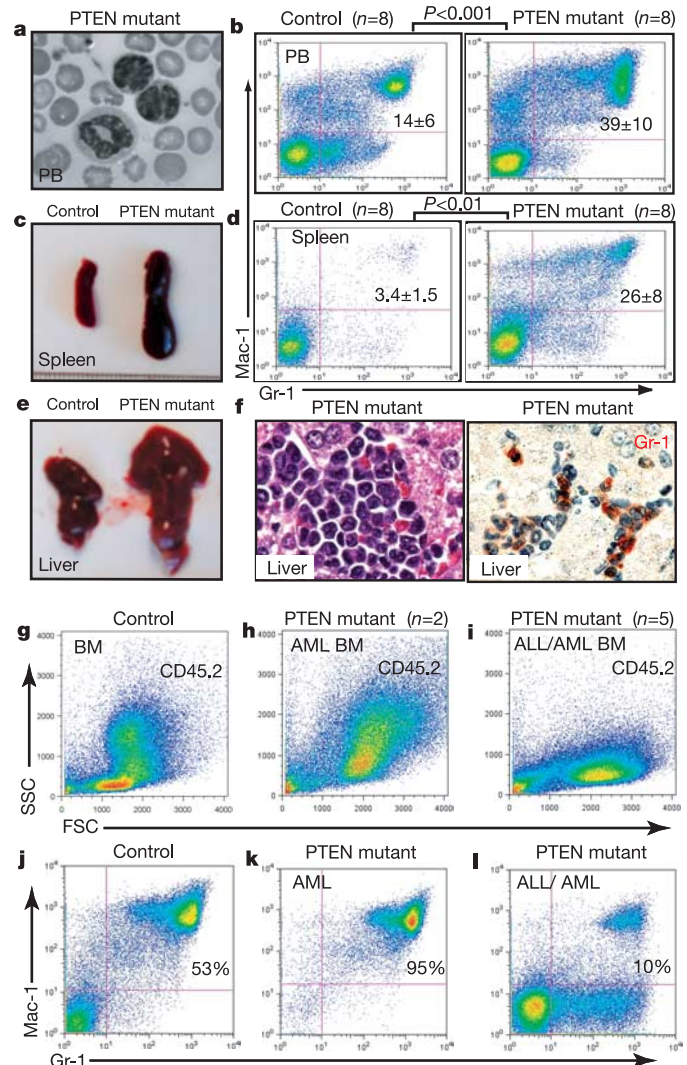


Figure 4 | Loss of PTEN results in myeloproliferative disorder (MPD) and leukaemia. **a–f**, Occurrence of MPD in PTEN-mutant mice as evidenced by an increased number of proliferative myeloid progenitors found in the peripheral blood (**a**, **b**), spleen (**c**, **d**) and liver (**e**, **f**), where myeloid cell infiltration in PTEN mutants is shown by Gr-1 staining. All tissues were studied at 30 DPI. **g–i**, Leukaemia developed from transplantation of PTEN-mutant bone marrow. Flow cytometry analyses of bone marrow from lethally irradiated mice transplanted with control bone marrow (**g**, **j**) or PTEN-mutant bone marrow (**h**, **i**, **k**, **l**). Relative heterogeneity (**g**; control) or homogeneity (**h**, **i**; PTEN-mutant) of CD45.2⁺ donor cell types in the bone marrow, revealed by separating cells according to their light-scattering properties (SSC, side scatter; FSC, forward scatter). Further characterization of the bone marrow according to Gr-1 and Mac-1 marker expression in CD45.2⁺ donor cells from control (**j**) or PTEN-mutant (**k**, **l**) bone marrow.

and a developmental block of cells in the prepro-B-lymphoid (CD19⁺CD43⁺) stage (Fig. 3d). Similar results were obtained from competitive repopulation assays (described below), with a decrease in the B lineage but increases in the myeloid and T lineages (Fig. 3e).

To test the reconstitution ability of PTEN-deficient HSCs, we performed competitive repopulation assays that compare the contribution of marked (CD45.2⁺) donor cells to that of recipient cells (CD45.1⁺) following transplantation into lethally irradiated mice. The contribution (per cent CD45.2⁺) of donor-derived cells, analysed in the peripheral blood at the indicated times post-transplantation, progressively decreased for PTEN-mutant donors but not for control donors (Fig. 3e). Analysis of the bone marrow at the 12-week time point revealed that—despite using five times more mutant cells—the proportion of donor-derived LSK cells in the group transplanted with PTEN-mutant cells was decreased to one-eighth that of the control-transplanted group (Fig. 3f). Thus, the general decline observed in the contribution of PTEN-mutant cells to the bone marrow reflects a depletion of the PTEN-deficient HSCs. Similar results were obtained from transplanted spleen cells (Supplementary Fig. 4), indicating that the HSCs aberrantly located in the spleen of PTEN-mutant donor animals were functionally similar to those found in the bone marrow.

Loss of PTEN function also results in myeloproliferative disorder (MPD), which, in humans, will (in many cases) transform into leukaemia at later stages⁶. PTEN mutants at around 30 DPI showed increased populations of proliferative monocytes and granulocytes in the peripheral blood (Fig. 4a, b), spleen (Fig. 4c, d) and multiple tissues including liver (Fig. 4e, f). These are all features characteristic of MPD⁶. In our PTEN mutants, the MPD leads to the death of the animals at around 30–40 DPI, so we performed bone marrow transplants to extend the time over which the impact of PTEN loss could be studied. Mice transplanted with PTEN-mutant bone marrow cells (which were combined with normal rescue bone marrow cells necessary to sustain the recipient) showed the same features of MPD that were observed in primary (donor) PTEN-mutant mice, but also subsequently developed severe leukaemia or lymphoma and died within 3–4 months. Mice that received control bone marrow showed a normal heterogeneity of cell types (Fig. 4g, j), whereas the profiles obtained from recipients of PTEN-mutant bone marrow (Fig. 4h, i, k, l) were more homogeneous, reflecting a blast crisis stage of acute myeloid leukaemia (AML; 2 out of 7 cases) (Fig. 4h, k) or acute myeloid and lymphoid leukaemia (AML/ALL; 5 out of 7 cases) (Fig. 4i, l). Two of the five AML/ALL mice progressed further and developed T lymphoma (data not shown). In mice with AML, more than 80% of the bone marrow cells expressed the myeloid Gr-1/Mac-1 markers (Fig. 4k). Mice that developed T-lymphoid leukaemia showed one of two phenotypes. In three out of five cases, the bone marrow was dominated by CD3⁺/CD4⁺T lymphocytes (>80%; Supplementary Fig. 5a). In the other two cases, over 80% of bone marrow cells were CD44⁺/CD25⁺, indicative of an enriched population of pro-T cells (Supplementary Fig. 5b), and only 10–20% of cells were myeloid (Fig. 4l).

In adults, homeostasis between the quiescent and activated states of stem cells is essential to balance stem cell maintenance with ongoing tissue regeneration^{1,3}. This study has shown that inactivation of PTEN deregulates HSCs, reducing the proportion that are quiescent (in G0), increasing the proportion that actively cycle, and compromising the ability to maintain sufficient LT-HSCs. We suggest that PTEN is a critical component of a molecular switch governing the forward (G0–G1) or backward (G1–G0) transitions between the quiescent and activated states in LT-HSCs—a model that is consistent with the impact of PTEN loss on HSCs and with our observation that PTEN in LT-HSCs is primarily associated with the cyclin D1⁺ population. In the absence of PTEN, cell cycle entry (the forward transition) may occur more readily, as recently suggested for neural stem cells²¹, or the switch may become ‘forward only’ and prevent LT-HSCs in G1 from returning to a quiescent state. The distinct

populations of LT-HSCs containing p-PTEN and non-p-PTEN raises the possibility that PTEN phosphorylation may serve as a sensor for this switch. Depending on environmental cues, the activated (G1) LT-HSCs may progress further through the cell cycle—a process controlled, in part, by PI(3)K–Akt activity. A potential downstream target at the G1-to-S transition is Myc activity, which can be controlled by Akt through phosphorylation of GSK3 β (ref. 22). PI(3)K–Akt also counters the cell cycle inhibitors p21 and p27 (ref. 4), so it is notable that the constitutive activation of HSCs we obtained by inactivation of PTEN resembles the result of deleting p21 in this population²³. Equally, leukaemia in animals carrying the Akt mutation can be partially rescued by treatment with rapamycin, an inhibitor of the serine–threonine kinase mTOR (mammalian target of rapamycin)²⁴. The forward and reverse transitions involve changes not only in cell cycle status but also in niche interaction properties²⁵, so the PTEN–Akt pathway, which affects both of these components, may help coordinate the two. The precise molecular basis of the aberrant mobilization of HSCs that have a PTEN mutation to the peripheral systems remains to be determined, but PTEN-mutant cells appear to be deficient in their retention ability in bone marrow niches.

PTEN also has a role in controlling haematopoietic lineage fate, as evidenced by the decline in CLP and B-lineage numbers and the increase in the myeloid and T lineages that was observed in PTEN mutants. PTEN deficiency could distort multiple lineages if PTEN governs lineage choice in the lymphoid primed multipotent progenitor (LMPP) that is proposed to give rise to the T lineage, CLPs and GMPs^{26–28} (see the model in Supplementary Fig. 5c). Most likely, however, PTEN acts at multiple stages as implied by lineage-specific PTEN inactivation studies^{29,30}. Inactivation of PTEN results in transplantable MPD and acute myeloid/lymphoid leukaemias that express markers of the myeloid and T lineages. These are strikingly similar to the expanded cell populations observed in transplanted mice with bone marrow cells carrying a PTEN mutation, indicative of an intrinsic link between lineage regulation and leukaemogenesis. This may be because constitutive PI(3)K–Akt activity distorts lineage commitment and provides a growth advantage for certain progenitor cells over other lineages, resulting in expanded cell pools in which a second genetic mutation may occur. Progenitor pool expansion in leukaemia contrasts with the behaviour of PTEN-deficient HSCs, which undergo uncontrolled proliferation but eventually become depleted or exhausted. Exploring this difference between leukaemic cells and HSCs should provide insight into self-renewal mechanisms, and could identify potential therapeutic targets for eliminating leukaemic stem cells without adversely affecting normal stem cells.

METHODS

Additional details are provided in Supplementary Methods.

PTEN-mutant mice. *Pten*^{fl/fl} mice⁹ were crossed with *Pten*^{fl/+} animals carrying an *Mx-1-Cre* transgene⁸ to generate litters containing *Mx-1-Cre*⁺; *Pten*^{fl/fl}, *Mx-1-Cre*⁺; *Pten*^{fl/+}, *Pten*^{fl/fl} and *Pten*^{fl/+} mice. PlpC injection at weaning induced PTEN deletion.

Flow cytometry. The isolation and preparation of bone marrow, spleen and peripheral blood cells, and the method for subsequent flow cytometry assays, have been described previously⁷.

Immunofluorescence studies. The immunostaining procedures used in this study have been described previously⁷.

Colony-forming unit assays. See Supplementary Methods.

Cell migration and adhesion assays. Adhesion molecules were analysed by flow cytometry as described previously¹⁷. For details on the cell migration assays, see Supplementary Methods.

Cell proliferation and cell cycle analysis. Cell proliferation and cell cycle profiles were assessed by flow cytometry on the basis of Ki67, pyronin Y and Hoechst 33342 staining.

Competitive engraftment assay and leukaemia phenotypes. Bone marrow cells from control and PTEN-mutant mice (CD45.2⁺ background) were each mixed with recipient bone marrow cells (CD45.1⁺ background) and transplanted into lethally irradiated recipient mice. At different time points post-transplantation, the recipient mice were analysed for the reconstitution ability of donor cells

using different lineage markers. Tissues from the recipient mice that developed acute leukaemia were collected for further cell-surface marker staining.

Homing and lodging assays. Bone marrow cells were obtained from control and PTEN-mutant mice and labelled with carboxyfluorescein diacetate, succinimidyl ester (CFDA-SE) (Invitrogen) as per the manufacturer's instructions. The homing and lodging assays involved transplanting labelled bone marrow cells into lethally irradiated and non-irradiated mice, respectively. At different time points, recipient-derived bone marrow and spleen cells were evaluated by flow cytometry for donor cells labelled with CFDA.

Received 6 October 2005; accepted 20 March 2006.

Published online 23 April 2006.

- Li, L. & Xie, T. Stem cell niche: structure and function. *Annu. Rev. Cell Dev. Biol.* **21**, 605–631 (2005).
- Heissig, B. *et al.* Recruitment of stem and progenitor cells from the bone marrow niche requires MMP-9 mediated release of kit-ligand. *Cell* **109**, 625–637 (2002).
- Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149–161 (2004).
- Stiles, B., Groszer, M., Wang, S., Jiao, J. & Wu, H. PTENless means more. *Dev. Biol.* **273**, 175–184 (2004).
- Cully, M., You, H., Levine, A. J. & Mak, T. W. Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis. *Nature Rev. Cancer* **6**, 184–192 (2006).
- Kogan, S. C. *et al.* Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice. *Blood* **100**, 238–245 (2002).
- Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
- Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269**, 1427–1429 (1995).
- Lesche, R. *et al.* Cre/loxP-mediated inactivation of the murine *Pten* tumor suppressor gene. *Genesis* **32**, 148–149 (2002).
- Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl Acad. Sci. USA* **98**, 14541–14546 (2001).
- Vazquez, F. *et al.* Phosphorylation of the PTEN tail acts as an inhibitory switch by preventing its recruitment into a protein complex. *J. Biol. Chem.* **276**, 48627–48630 (2001).
- He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt- β -catenin signaling. *Nature Genet.* **36**, 1117–1121 (2004).
- Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
- Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Lapidot, T., Dar, A. & Kollet, O. How do stem cells find their way home? *Blood* **106**, 1901–1910 (2005).
- Adams, G. B. *et al.* Stem cell engraftment at the endosteal niche is specified by the calcium-sensing receptor. *Nature* **439**, 599–603 (2006).
- van der Loo, J. C. *et al.* VLA-5 is expressed by mouse and human long-term repopulating hematopoietic cells and mediates adhesion to extracellular matrix protein fibronectin. *J. Clin. Invest.* **102**, 1051–1061 (1998).
- Pasgue, E., Wagers, A. J., Giuriato, S., Anderson, W. C. & Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *J. Exp. Med.* **202**, 1599–1611 (2005).
- Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
- Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
- Groszer, M. *et al.* PTEN negatively regulates neural stem cell self-renewal by modulating G0–G1 cell cycle entry. *Proc. Natl Acad. Sci. USA* **103**, 111–116 (2006).
- Gregory, M. A., Qi, Y. & Hann, S. R. Phosphorylation by glycogen synthase kinase-3 controls c-myc proteolysis and subnuclear localization. *J. Biol. Chem.* **278**, 51606–51612 (2003).
- Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21^{cip1/waf1}. *Science* **287**, 1804–1808 (2000).
- Wendel, H. G. *et al.* Survival signalling by Akt and eIF4E in oncogenesis and cancer therapy. *Nature* **428**, 332–337 (2004).
- Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–2763 (2004).
- Adolfsson, J. *et al.* Identification of Flt3⁺ lympho-myeloid stem cells lacking erythro-megakaryocytic potential. A revised road map for adult blood lineage commitment. *Cell* **121**, 295–306 (2005).
- Igarashi, H., Gregory, S. C., Yokota, T., Sakaguchi, N. & Kincade, P. W. Transcription from the *RAG1* locus marks the earliest lymphocyte progenitors in bone marrow. *Immunity* **17**, 117–130 (2002).
- Akashi, K., Traver, D. & Zon, L. I. The complex cartography of stem cell commitment. *Cell* **121**, 160–162 (2005).
- Suzuki, A. *et al.* T cell-specific loss of Pten leads to defects in central and peripheral tolerance. *Immunity* **14**, 523–534 (2001).
- Suzuki, A. *et al.* Critical roles of Pten in B cell homeostasis and immunoglobulin class switch recombination. *J. Exp. Med.* **197**, 657–667 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Sherr, Y. He and P. Kincade for scientific discussion, and D. di Natale for manuscript proofreading. We are grateful to L. Shannon for assistance in genotyping; J. Wunderlich for flow cytometry assistance; T. Johnson and her co-workers for histology assistance; and H. Marshall for technical assistance. This work was mainly supported by the Stowers Institute for Medical Research.

Author Contributions J.Z., X.C.H., J.S.H., L.M.W. and L.L. designed the research; J.Z., X.C.H., T.Y., J.S.H. and S.J. performed the research; J.T.R., D.R., J.S.H., S.J. and K.S.P.-W. provided technical support; H.W. contributed critical reagents; J.Z., X.C.H., J.C.G., J.S.H., S.J., L.M.W. and L.L. analysed data; and L.L., L.M.W. and J.C.G. wrote the paper.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.L. (lil@stowers-institute.org).

Identification of a tumour suppressor network opposing nuclear Akt function

Lloyd C. Trotman^{1,2}, Andrea Alimonti^{1,2}, Pier Paolo Scaglioni^{1,3}, Jason A. Koutcher⁴, Carlos Cordon-Cardo² & Pier Paolo Pandolfi^{1,2}

The proto-oncogene AKT (also known as PKB) is activated in many human cancers, mostly owing to loss of the *PTEN* tumour suppressor¹. In such tumours, AKT becomes enriched at cell membranes where it is activated by phosphorylation. Yet many targets inhibited by phosphorylated AKT (for example, the FOXO transcription factors) are nuclear; it has remained unclear how relevant nuclear phosphorylated AKT (pAKT) function is for tumorigenesis. Here we show that the PML tumour suppressor prevents cancer by inactivating pAKT inside the nucleus. We find in a mouse model that *Pml* loss markedly accelerates tumour onset, incidence and progression in *Pten*-heterozygous mutants, and leads to female sterility with features that recapitulate the phenotype of *Foxo3a* knockout mice². We show that *Pml* deficiency on its own leads to tumorigenesis in the prostate, a tissue that is exquisitely sensitive to pAkt levels, and demonstrate that *Pml* specifically recruits the Akt phosphatase PP2a as well as pAkt into Pml nuclear bodies. Notably, we find that *Pml*-null cells are impaired in PP2a phosphatase activity towards Akt, and thus accumulate nuclear pAkt. As a consequence, the progressive reduction in *Pml* dose leads to inactivation of *Foxo3a*-mediated transcription of proapoptotic *Bim* and the cell cycle inhibitor *p27^{kip1}*. Our results demonstrate that Pml orchestrates a nuclear tumour suppressor network for inactivation of nuclear pAkt, and thus highlight the importance of AKT compartmentalization in human cancer pathogenesis and treatment.

Pml nuclear bodies represent distinct yet dynamic intranuclear structures involved in apoptosis, proliferation and senescence³. Nuclear bodies are absent from *Pml*-deficient cells, causing aberrant nuclear patterns of nuclear body resident proteins. We have recently reported the high correlation between reduction in nuclear body number and prostate and colon cancer⁴, both tumours showing frequent *PTEN* loss^{5,6}. Because we and others have established faithful mouse models for the role of *PTEN* in cancer^{7–9}, we sought to determine whether *Pml* loss would affect the signature of *Pten* deficiency. After crossing *Pten*-heterozygous mice (*Pten*^{+/-}) with *Pml*-null mice (*Pml*^{-/-}) to produce six genotypes of interest, we found a marked *Pml*-dependent reduction in lifespan of *Pten*^{+/-} mice (Supplementary Fig. S1a). *Pten*^{+/-} female lethality, dictated by an autoimmune disorder¹⁰, was not precipitated by *Pml* loss. In contrast, the male cohort displayed an exquisite, *Pml* dose dependence in survival (the autoimmune disorder is minor in males¹⁰), thus exposing a novel lethal phenotype in compound mutants with *Pml* haploinsufficiency for its repression. Magnetic resonance imaging (MRI) visualized intestinal and prostate anomalies (not shown), and post-mortem analysis confirmed large intestinal lesions in *Pten*^{+/-} *Pml*^{+/-} and *Pten*^{+/-} *Pml*^{-/-} mice, suggesting obstruction as a probable (males) or additional (females) cause of death. Intestines of *Pten*^{+/-} *Pml*^{-/-} mice presented invasive adenocarci-

noma of the colon (Fig. 1a, colon), whereas *Pten*^{+/-} mice displayed only pre-cancerous polyps (Supplementary Fig. S1b, colon). Disease-free survival analysis summarizes onset of the *Pten*^{+/-} *Pml*^{-/-}-specific colon carcinoma (Fig. 1b). Notably, *Pml* levels also dictated polyp numbers in *Pten*^{+/-} animals: *Pml* dose reduction resulted in a 3–5-fold increase in the number of polyps per mouse (Fig. 1b, top inset). Average polyp diameter also correlated with *Pml* status (Fig. 1b, bottom inset). Taken together, in *Pten*^{+/-} mice, *Pml* loss causes the appearance of polyposis and colon cancer, which reduces lifespan severely.

Pten activity regulates prostate cancer, as by 8 months some *Pten*^{+/-} animals develop prostatic intraepithelial neoplasia (the *in situ* form of prostate cancer). Invasive cancer is observed on further lowering of *Pten* activity, as shown in *Pten*-hypomorphic or prostate-specific null mice^{11,12}. Haematoxylin and eosin staining and prostate cancer-free survival analysis revealed that *Pten*^{+/-} *Pml*^{-/-} and *Pten*^{+/-} *Pml*^{+/-} males (but not *Pten*^{+/-} mice) developed highly invasive cancers, suggesting that prostate epithelia are more sensitive to *Pml* status than intestine (Fig. 1 and Supplementary Fig. S1b, prostate). We also found a characteristic increased cell proliferation marker staining¹³ (even in tumour-free areas) in *Pten*^{+/-} *Pml*^{+/-} and *Pten*^{+/-} *Pml*^{-/-} intestines as well as in *Pten*^{+/-} *Pml*^{-/-} prostate glands (Supplementary Fig. S1c).

The degree of *Pten* deficiency closely correlates with activation of the oncogenic kinase Akt¹¹. Akt activation by membrane recruitment is inhibited by *Pten*^{14,15} and leads to phosphorylation of Akt at Thr 308 and Ser 473 (refs 16, 17). To address whether *Pml* loss affects Akt activation we performed immunohistochemistry (IHC) staining and quantified western blotting of tissues. Staining of pAkt at Ser 473 was clearly increased in polyps, normal colon and in prostate lesions of *Pten*^{+/-} *Pml*^{-/-} mice when compared with *Pten*^{+/-} mice (Fig. 2a). Importantly, both colon and prostate presented dominant nuclear pAkt accumulation, indicating a qualitative change in kinase localization (Fig. 2a, insets). Quantification of pAkt levels in *Pten*^{+/-} mice confirmed a 50% increase in normal colon and 2-month-old prostate on *Pml* loss (Supplementary Fig. S1d, compare *Pten*^{+/-} *Pml*^{-/-} (en) and *Pten*^{+/-} *Pml*^{+/-} (ew)), comparable to the reported *Pten*^{+/+} to *Pten*^{+/-} transition¹¹. Loss of the remaining *Pten* allele was excluded through western blot (not shown), IHC and Southern blot analysis of the macro-dissected cancers (Supplementary Fig. S1e). Taken together, our data suggest that in *Pten*^{+/-} mice, loss of *Pml* leads to further activation of Akt but not because of complete *Pten* loss.

When analysing *Pten*^{+/+} prostates we found that *Pml*^{-/-} animals presented high-grade prostatic intraepithelial neoplasia and areas of focally invasive cancer in their anterior prostates around 12 months of age (Fig. 2b and Supplementary Fig. S2a, b). Normal colon and prostate in *Pml*-deficient mice showed increased levels of pAkt

¹Cancer Biology and Genetics Program, ²Department of Pathology, ³Department of Medicine, ⁴Departments of Medicine, Radiology and Medical Physics, Memorial Sloan-Kettering Cancer Center, Sloan-Kettering Institute, 1275 York Avenue, New York, New York 10021, USA.

(Fig. 2c) but no signs of colon neoplasia/dysplasia, consistent with the above finding that prostate is most sensitive to *Pml* loss (Fig. 1b). We frequently found enlarged prostates in *Pml*-null mice, yet larger organ size was also due to large luminal non-cellular areas (see Supplementary Fig. S2c, asterisk in the *Pten*^{+/-} *Pml*^{+/-} panel). To measure size effects we determined average individual cell (not organ) sizes. *Pml*^{-/-} and *Pten*^{+/-} glands were mostly hypercellular but filled with cells of only half the size of their wild-type counterparts (Supplementary Fig. S2c). Thus, *Pml* deficiency leads to Akt activation in prostate and colon and triggers initiation of prostate cancer.

To test whether the qualitative difference in pAkt localization was an inherent feature of *Pml* loss we used mouse embryonic fibroblasts (MEFs). First, we determined Akt status in primary littermate *Pml*^{+/-} and *Pml*^{-/-} MEFs by subjecting them to serum starvation and

stimulation (see Methods). These cells had little and comparable pAkt activation at steady state and during serum starvation (Supplementary Fig. S2d, top panels). But serum stimulation for 10 min gave stronger Akt activation in *Pml*^{-/-} than in *Pml*^{+/-} MEFs, as confirmed by Akt kinase assays (Supplementary Fig. S2e). At 3 h after stimulation, Akt activation was slightly higher in null MEFs, an effect more readily observed at later passage (Supplementary Fig. S2f, top panels). To test whether transgenic expression of human PML also affects Akt activation we used stably PML-transduced *Pml*^{-/-} cells, which showed decreased Akt activation and faster inactivation after stimulation (Supplementary Fig. S2d, bottom panels). PML overexpression in *Pml*^{+/-} MEFs had almost no effect (not shown), similar to overexpression in PML^{+/-} human cells (Supplementary Fig. S2f, middle panels). Thus, Pml is able to suppress Akt activation after a membrane stimulus and it is sufficient to remedy this defect in *Pml*^{-/-} cells. To exclude effects on Akt caused by cell cycle perturbations after PML add back, cell cycle profiles of *Pml*^{-/-} and add-back cells were compared but no differences found, especially after starvation/stimulation (Supplementary Fig. S2g).

To account for qualitative and quantitative differences by confocal laser scanning microscopy (CLSM), primary *Pml*^{+/-} and *Pml*^{-/-} MEFs were seeded on the same coverslips. Minor differences in pAkt staining were observed between wild-type and null cells at steady state and starvation (Fig. 3a). In contrast, *Pml*^{-/-} cells showed not only stronger but notably nuclear pAkt accumulation after 10 min or 3 h of serum stimulation (see also Supplementary Fig. S3a). This localization was in marked contrast to that observed in *Pten*^{-/-} cells, where pAkt is concentrated at cell membranes (Fig. 3a, bottom row). This membrane-associated pAkt pool was not antagonized by overexpression of PML in these MEFs (not shown), or in human PTEN^{-/-} cells (Supplementary Fig. S2f, bottom panels), consistent with inhibition of nuclear, not cytoplasmic, pAkt by PML. Cell fractionation confirmed that a distinct nuclear population of pAkt was present in *Pml*^{-/-} but not *Pml*^{+/-} cells after stimulation (Supplementary Fig. S3b). We next compared pAkt localization in three invasive prostate cancer models. The prostate-specific *Pten*-null

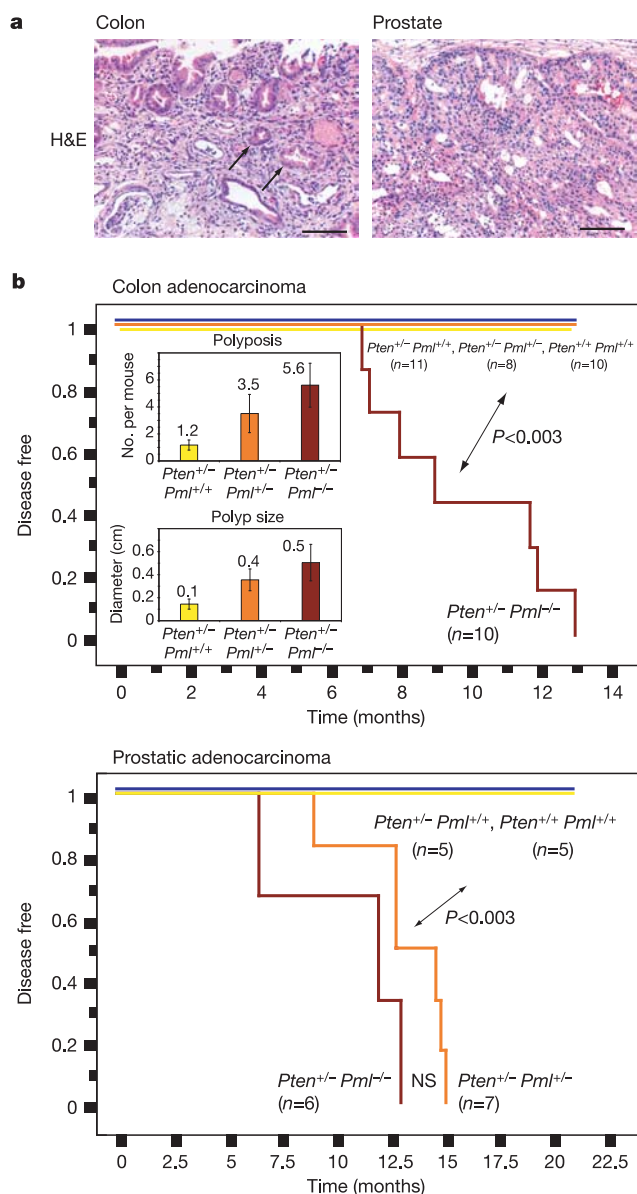


Figure 1 | *Pml* status dictates carcinogenesis in *Pten*^{+/-} mice.

a, Haematoxylin and eosin (H&E) analysis of colon and prostate tissue reveals invasive adenocarcinoma in *Pten*^{+/-} *Pml*^{-/-} mice. Arrows show invasive glands in colonic submucosa. Scale bars, 100 μ m. **b**, Kaplan-Meier plots for disease-free survival. Average polyps per mouse (top inset) and average polyp size (bottom inset) are shown. Error bars are s.d. (see also Methods).

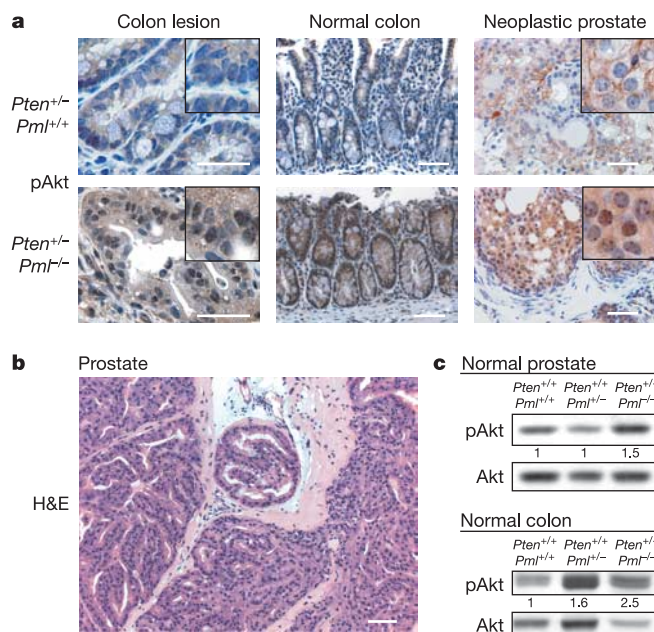


Figure 2 | *Pml* loss leads to an increase in pAkt *in vivo*. **a**, IHC staining comparing pAkt levels and localization in *Pten*^{+/-} *Pml*^{+/-} and *Pten*^{+/-} *Pml*^{-/-} mice. **b**, Haematoxylin and eosin (H&E) staining of *Pml*^{-/-} prostate with high-grade prostatic intraepithelial neoplasia. Scale bars (**a**, **b**), 100 μ m. **c**, *Pml* status affects pAkt levels in pre-neoplastic *Pten*^{+/-} tissues.

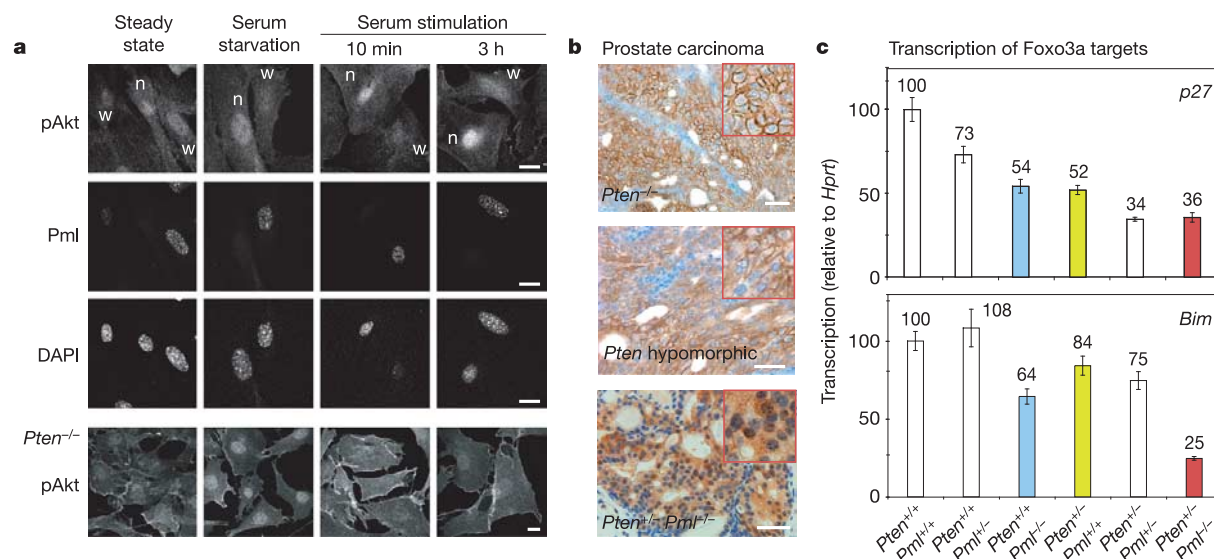


Figure 3 | *Pml* deficiency leads to increased nuclear pAkt localization and function *in vitro* and *in vivo*. **a**, Immunofluorescence CLSM on primary littermate *Pml*^{+/+} (w) and *Pml*^{-/-} (n) MEFs seeded on the same coverslips. pAkt in *Pten*^{-/-} MEFs is shown for comparison. Scale bars, 10 μ m. **b**, pAkt

in immunohistochemistry staining of prostate cancers from indicated mouse models. Scale bars, 50 μ m. **c**, Foxo3a activity measured by quantitative real-time PCR of prostates of the indicated genotypes. Error bars are s.d. of triplicates.

and *Pten*-hypomorphic models showed strong membrane accumulation of pAkt, as previously described¹¹ (Fig. 3b). In contrast, the invasive prostate cancer of *Pten*^{+/+} *Pml*^{-/-} mice showed dominant nuclear and cytoplasmic staining (see also Fig. 2a, insets). Therefore, we next tested whether *Pml* loss would enhance nuclear pAkt function.

The forkhead transcription factor FOXO3A is a well-studied nuclear target of pAKT¹⁸, as AKT-mediated phosphorylation causes FOXO3A inactivation and export¹⁹. We quantified nuclear Foxo3a amounts by western blotting and found that *Pml*^{-/-} cells, in contrast to wild-type cells, had decreased exogenous and endogenous nuclear Foxo3a after serum stimulation (Supplementary Fig. S3c). To quantify Foxo3a distribution by non-disruptive means, GFP-FOXO3A (human FOXO3A tagged with green fluorescent protein) localization

in fixed *Pml*^{+/+} and *Pml*^{-/-} MEFs was scored. At 10 min after stimulation, 80% of *Pml*^{-/-} cells showed cytoplasmic FOXO3A whereas 70% of wild-type cells still retained FOXO3A in the nucleus (Supplementary Fig. S3d). Importantly, an Akt-insensitive FOXO3A mutant retained its nuclear localization even after a 3 h stimulation in *Pml*^{-/-} cells. In agreement, a twofold increase in Akt-mediated FOXO3A phosphorylation was observed after serum stimulation when comparing *Pml*-deficient cells with add-back cells (Supplementary Fig. S3e).

Foxo3a exerts some of its tumour-suppressive functions by inducing transcription of *p27^{kip1}* (see ref. 18), which inhibits prostate cancer after *Pten* loss¹³. As quantified by real-time polymerase chain reaction (PCR), *Pml* status effectively dictates *p27^{kip1}* messenger RNA levels in prostates of both *Pten*^{+/+} and *Pten*^{+/+} mice (Fig. 3c).

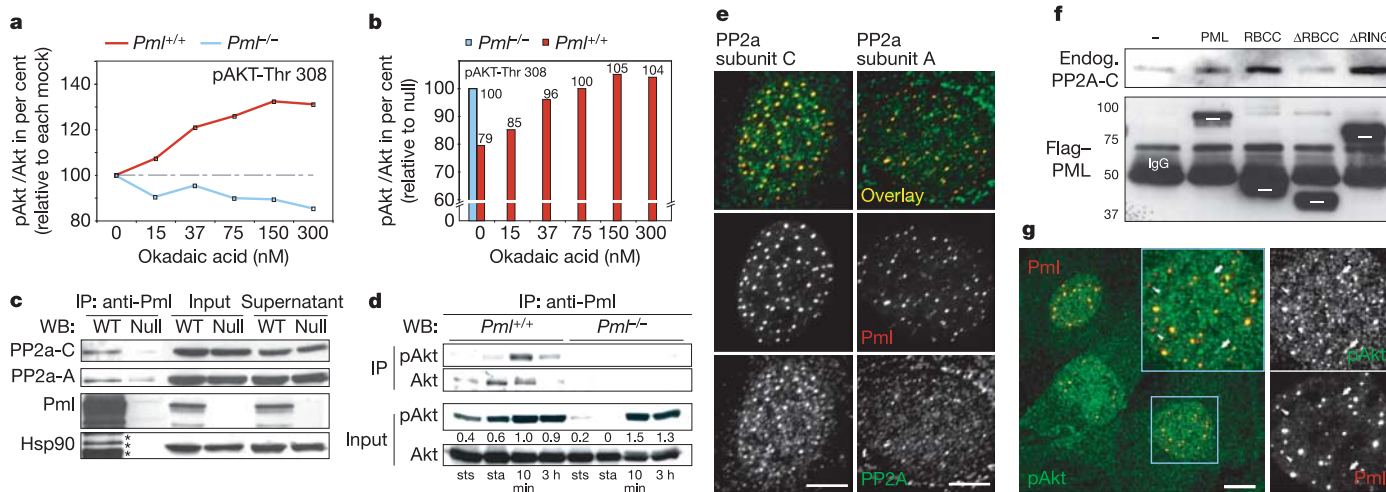


Figure 4 | *Pml* affects PP2a activity towards pAkt. **a**, pAkt-Thr 308/Akt ratios after okadaic acid treatment of MEFs (relative to mock, see Methods). **b**, Ratios from **a** expressed relative to untreated *Pml*^{-/-} cells. Note the broken ordinate axis for clarity. **c**, Endogenous co-immunoprecipitation of PP2a with Pml in MEFs. Asterisks indicate leftover Pml staining, not Hsp90. **d**, Endogenous co-immunoprecipitation of Akt with Pml at steady state (sts), starvation (sta) and serum stimulation for times indicated.

e, CLSM co-localization of endogenous Pml and PP2a in MEFs. Scale bars, 5 μ m. **f**, Anti-Flag co-immunoprecipitation of PP2a-C with Flag-tagged PML/PML mutants. White bars indicate PML/mutant migration. IgG, anti-Flag IgG heavy chains. Molecular weights are in kDa. **g**, CLSM co-localization of pAkt and Pml. Arrowheads indicate pAkt-deficient nuclear bodies and arrows indicate pAkt accumulation outside nuclear bodies. Scale bar, 10 μ m.

Two additional Foxo3a target genes—the proapoptotic factor Bim and DNA repair protein Gadd45 (ref. 18)—responded similarly, consistent with Foxo3a inactivation (see also Supplementary Fig. S3f).

Foxo3a-deficient females display early infertility due to premature ovarian failure^{20,21}, an event normally associated with ageing². To determine the effect of *Pml* loss on fertility we set up mating pairs using fertile wild-type males. *Pml*^{-/-} and *Pten*^{+/-} females accumulated litters (Supplementary Fig. S4a) similar to wild-type mice (not shown). In contrast, although *Pten*^{+/-} *Pml*^{-/-} females had near-normal first litter sizes, they were invariably infertile by 5 months of age and their ovaries showed signs of follicular degeneration at 2 months and near-complete follicle depletion by 5 months (Supplementary Fig. S4b). Taken together, our results demonstrate that *Pml* loss enhances Foxo3a inactivation, especially under conditions of elevated Akt (*Pten* loss *in vivo*, serum stimulation *in vitro*).

Mechanistically, *Pml* loss could either stimulate Akt activation or antagonize pAkt inactivation. To distinguish between the two, we used okadaic acid, which selectively inactivates PP2a, the only known Akt-Thr 308 phosphatase^{22,23}. If *Pml* deficiency favours Akt activation mechanisms (independent of PP2a), okadaic acid treatment should increase Akt phosphorylation at Thr 308 in *Pml*^{-/-} cells. Alternatively, if *Pml* deficiency antagonizes PP2a, *Pml*^{-/-} cells should be insensitive to okadaic acid. Our results suggest the latter: *Pml*^{-/-} MEFs, in contrast to wild type, showed no response in the levels of pAkt at Thr 308 on okadaic acid addition in fresh serum (Fig. 4a and Supplementary Fig. S4c). Thus, after this stimulation in the presence of okadaic acid, *Pml*^{+/-} cells were finally able to match the pAkt levels of *Pml*^{-/-} cells (Fig. 4b). Importantly, *Pml*^{-/-} cell extracts retained normal PP2a levels and general activity, suggesting that Pml specifically cooperates with PP2a in pAkt inactivation (Supplementary Fig. S4d, e).

PP2a is a heterotrimer of A, B and catalytic (C) subunits, which can function alone or in a 'core dimer' with A. Whereas the two known A and C subunit isoforms are ubiquitous, the B subunit families can determine tissue/substrate specificity and give rise to over 70 holoenzyme variations²⁴. We found the C and to some extent the A subunits co-precipitating with Pml (Fig. 4c), which prompted us to test whether PP2a could act in nuclear bodies. Co-immunoprecipitation experiments showed that Akt is able to associate with Pml both when activated or not (Fig. 4d); using CLSM of primary MEFs, significant nuclear body co-localization of PP2a-C and PP2a-A with Pml as well as nuclear body enrichment was evident (Fig. 4e), and PP2a-C was also found enriched in many *Pml*^{+/-} (but not *Pml*^{-/-}) nuclei (Supplementary Fig. S4f). No Pml co-localization was found in the randomly tested PP2a-B PR-55 family members (Supplementary Fig. S4g).

As a result of probing PML deletion mutants for their ability to bind endogenous PP2a-C, we found that deletion of the RBCC motif abolished binding whereas the RING finger motif was dispensable (Fig. 4f), thus delineating the necessary interaction domains (Supplementary Fig. S5a). Similarly, Δ RBCC was unable to antagonize pAkt activation (Supplementary Fig. S5b). PP2a might also affect Akt indirectly via PDK1 (ref. 25), but we found unaltered steady-state phosphorylated PDK1 levels in tissues and MEFs of various genotypes (not shown) and similar steady-state levels of pAkt in *Pml*^{+/-} and *Pml*^{-/-} cells (Supplementary Figs S3b, cytoplasm, and S5c, nucleus). Specific enrichment of pAkt in Pml nuclear bodies was found in wild-type MEFs when visualizing subnuclear pAkt localization by CLSM (Fig. 4g). Taken together, our data suggest that Pml can markedly increase effective concentrations of PP2a and its target, pAkt, in nuclear bodies, consistent with a nuclear pAkt clearing deficiency of *Pml*^{-/-} cells.

Characterization of the molecular pathways leading to cancer is a major step towards understanding and combating the disease²⁶. Here we have used mouse genetics to gather insights into AKT-driven

tumorigenesis and established a mouse model for epithelial cancers triggered by *Pml* loss.

First, we have demonstrated the importance of *Pml* gene dose in prostate and colon carcinoma especially after *Pten* loss. For comparison, *Trp53*^{-/-} prostates never show neoplasia²⁷, and *Pten*^{+/-} *Trp53*^{-/-} prostates (ref. 27) or *Pten*^{+/-} *pRb*^{+/-} mice show no invasive prostate cancer (M. Niki and P.P., unpublished data). Second, PML loss highlights nuclear AKT function, a novel quality in PTEN-loss-driven tumorigenesis for which no model system exists so far. Colon, male prostate glands and female ovaries appear especially sensitive to increases in nuclear Akt and decreased Foxo3a levels. As we have shown that *p27*^{kpi1} becomes a prostate and colon cancer inhibitor in conditions of *Pten* heterozygosity¹³, these results highlight the decisive role of FOXO3A function in human tumours with loss of one PTEN copy. Finally, we have identified a Pten–Pml–PP2a tumour-suppressive network for the inactivation of nuclear pAkt (see model, Supplementary Fig. S5d). Thus, PML regulates the activity of major opponents of AKT signalling in epithelia, which allows us to envision PML-stabilizing approaches for therapy of AKT-driven cancers.

METHODS

Mice. *Pml*^{-/-} (ref. 28) and *Pten*^{+/-} (ref. 7) mutants were crossed as described in Supplementary Information. The six cohort genotypes were: (1) *Pten*^{+/-} *Pml*^{+/-}; (2) *Pten*^{+/-} *Pml*^{+/-}; (3) *Pten*^{+/-} *Pml*^{-/-}; (4) *Pten*^{+/-} *Pml*^{+/-}; (5) *Pten*^{+/-} *Pml*^{+/-}; and, (6) *Pten*^{+/-} *Pml*^{-/-}.

MRI. For initial tumour assessment, mice of all genotypes were subjected to monthly MRI screening similar to the protocol described previously¹¹ (see also Supplementary Information).

Western blotting, immunoprecipitation and *in vitro* kinase assays. Tissue and cell lysates were prepared as previously described¹¹ and processed as outlined in Supplementary Information. *In vitro* kinase assays were done using a non-radioactive kinase kit (see Supplementary Information).

Cells, plasmids and immunofluorescence. *Pten*^{-/-} *Trp53*^{-/-} MEFs were prepared as described²⁷. Plasmids for FOXO3A and a FOXO3A mutant¹⁹ were gifts of M. Greenberg. The Flag–PML IV-derived plasmids were done as previously described²⁹. See Supplementary Information for further information.

IHC and quantifications. For IHC, tissues were fixed and embedded in paraffin according to standard procedures (see Supplementary Information) and determination of average prostate cell size was calculated as outlined in Supplementary Information.

Quantitative real time PCR. Tissue RNA was isolated using the Trizol method and quantitative real-time PCR performed as described³⁰. See Supplementary Information for all primer sets and detailed methods.

A detailed Methods section including antibodies used is available as online Supplementary Information.

Received 15 February; accepted 13 April 2006.

Published online 7 May 2006.

- Luo, J., Manning, B. D. & Cantley, L. C. Targeting the PI3K–Akt pathway in human cancer: rationale and promise. *Cancer Cell* **4**, 257–262 (2003).
- Brenkman, A. B. & Burgering, B. M. FoxO3a eggs on fertility and aging. *Trends Mol. Med.* **9**, 464–467 (2003).
- Salomoni, P. & Pandolfi, P. P. The role of PML in tumor suppression. *Cell* **108**, 165–170 (2002).
- Gurrieri, C. et al. Loss of the tumour suppressor PML in human cancers of multiple histologic origins. *J. Natl Cancer Inst.* **96**, 269–279 (2004).
- Goel, A. et al. Frequent inactivation of PTEN by promoter hypermethylation in microsatellite instability-high sporadic colorectal cancers. *Cancer Res.* **64**, 3014–3021 (2004).
- Porkka, K. P. & Visakorpi, T. Molecular mechanisms of prostate cancer. *Eur. Urol.* **45**, 683–691 (2004).
- Di Cristofano, A., Pesce, B., Cordon-Cardo, C. & Pandolfi, P. P. Pten is essential for embryonic development and tumour suppression. *Nature Genet.* **19**, 348–355 (1998).
- Podsypanina, K. et al. Mutation of Pten/Mmac1 in mice causes neoplasia in multiple organ systems. *Proc. Natl Acad. Sci. USA* **96**, 1563–1568 (1999).
- Suzuki, A. et al. High cancer susceptibility and embryonic lethality associated with mutation of the PTEN tumor suppressor gene in mice. *Curr. Biol.* **8**, 1169–1178 (1998).
- Di Cristofano, A. et al. Impaired Fas response and autoimmunity in Pten^{+/-} mice. *Science* **285**, 2122–2125 (1999).
- Trotman, L. C. et al. Pten dose dictates cancer progression in the prostate. *PLoS Biol.* **1**, E59 (2003).

12. Wang, S. *et al.* Prostate-specific deletion of the murine *Pten* tumor suppressor gene leads to metastatic prostate cancer. *Cancer Cell* **4**, 209–221 (2003).
13. Di Cristofano, A., De Acetis, M., Koff, A., Cordon-Cardo, C. & Pandolfi, P. P. *Pten* and *p27KIP1* cooperate in prostate cancer tumor suppression in the mouse. *Nature Genet.* **27**, 222–224 (2001).
14. Maehama, T., Taylor, G. S. & Dixon, J. E. PTEN and myotubularin: novel phosphoinositide phosphatases. *Annu. Rev. Biochem.* **70**, 247–279 (2001).
15. Stambolic, V. *et al.* Negative regulation of PKB/Akt-dependent cell survival by the tumor suppressor PTEN. *Cell* **95**, 29–39 (1998).
16. Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor–mTOR complex. *Science* **307**, 1098–1101 (2005).
17. Brazil, D. P., Yang, Z. Z. & Hemmings, B. A. Advances in protein kinase B signalling: AKTion on multiple fronts. *Trends Biochem. Sci.* **29**, 233–242 (2004).
18. Tran, H., Brunet, A., Griffith, E. C. & Greenberg, M. E. The many forks in FOXO's road. *Sci. STKE* **2003**, RE5 (2003).
19. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
20. Castrillon, D. H., Miao, L., Kollipara, R., Horner, J. W. & DePinho, R. A. Suppression of ovarian follicle activation in mice by the transcription factor Foxo3a. *Science* **301**, 215–218 (2003).
21. Hosaka, T. *et al.* Disruption of forkhead transcription factor (FOXO) family members in mice reveals their functional diversification. *Proc. Natl Acad. Sci. USA* **101**, 2975–2980 (2004).
22. Millward, T. A., Zolnierowicz, S. & Hemmings, B. A. Regulation of protein kinase cascades by protein phosphatase 2A. *Trends Biochem. Sci.* **24**, 186–191 (1999).
23. Gao, T., Furnari, F. & Newton, A. C. PHLPP: a phosphatase that directly dephosphorylates Akt, promotes apoptosis, and suppresses tumor growth. *Mol. Cell* **18**, 13–24 (2005).
24. Janssens, V., Goris, J. & Van Hoof, C. PP2A: the expected tumor suppressor. *Curr. Opin. Genet. Dev.* **15**, 34–41 (2005).
25. Li, L., Ren, C. H., Tahir, S. A., Ren, C. & Thompson, T. C. Caveolin-1 maintains activated Akt in prostate cancer cells through scaffolding domain binding site interactions with and inhibition of serine/threonine protein phosphatases PP1 and PP2A. *Mol. Cell. Biol.* **23**, 9389–9404 (2003).
26. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
27. Chen, Z. *et al.* Crucial role of p53-dependent cellular senescence in suppression of *Pten*-deficient tumorigenesis. *Nature* **436**, 725–730 (2005).
28. Wang, Z. G. *et al.* PML is essential for multiple apoptotic pathways. *Nature Genet.* **20**, 266–272 (1998).
29. Guo, A. *et al.* The function of PML in p53-dependent apoptosis. *Nature Cell Biol.* **2**, 730–736 (2000).
30. Brunet, A. *et al.* Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science* **303**, 2011–2015 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Bernardi, B. Carver, Z. Chen, S. Clohessy, T. Maeda and T. Yung for discussion, help with reagents and data analysis; W. Golden and M. S. Jiao for help with pathology analysis; and C. Le, M. Lupu and C. Matei for help with MR imaging. This work was supported by NIH grants to P.P.P. and P.P.S. as well as grants to J.A.K. and the Memorial Sloan-Kettering Cancer Center.

Author Contributions The experiments were conceived and designed by L.C.T., A.A., P.P.S., J.A.K., C.C.-C. and P.P.P. Experiments were performed by L.C.T., A.A., P.P.S. and J.A.K. Data were analysed by L.C.T., A.A., P.P.S., J.A.K., C.C.-C. and P.P.P. The paper was written by L.C.T. and P.P.P.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.P.P. (p-pandolfi@ski.mskcc.org).

LETTERS

GTP-dependent twisting of dynamin implicates constriction and tension in membrane fission

Aurélien Roux¹, Katherine Uyhazi¹, Adam Frost² & Pietro De Camilli¹

Dynamin, a crucial factor in endocytosis^{1–3}, is a member of a family of GTPases that participates in membrane fission^{4–6}. It was initially proposed to act as a machine that constricts and cuts the neck of nascent vesicles in a GTP-hydrolysis-dependent reaction^{4,5}, but subsequent studies suggested alternative models^{2,7,8}. Here we monitored the effect of nucleotides on dynamin-coated lipid tubules in real time. Addition of GTP, but not of GDP or GTP- γ S, resulted in twisting of the tubules and supercoiling, suggesting a rotatory movement of the helix turns relative to each other during GTP hydrolysis. Rotation was confirmed by the movement of beads attached to the tubules. Twisting activity produced a longitudinal tension that was released by tubule breakage when both ends of the tubule were anchored. Fission also occurred when dynamin and GTP were added to lipid tubules that had been generated from liposomes by the motor activity of kinesin on microtubules. No fission events were observed in the absence of longitudinal tension. These findings demonstrate a mechano-enzyme activity of dynamin in endocytosis, but also imply that constriction is not sufficient for fission. At the short necks of endocytic vesicles, other factors^{6,9,10} leading to tension may cooperate with the constricting activity of dynamin to induce fission^{11–13}.

Purified dynamin has been shown, by negative-staining electron microscopy (EM), to tubulate lipid membranes and to fragment them into small vesicles on GTP hydrolysis^{14–16}. However, when the products of these incubations were frozen in suspension and analysed by cryo-EM, GTP-dependent constriction—but not fission—was observed¹⁷. It was proposed that tubule fragmentation observed in other assays might be the result of mechanical stress of the constricted tubules during sample preparation¹⁷. EM methods rely on static observations. To overcome this problem, we developed light-microscopy-based assays to monitor dynamically the effect of dynamin on lipid membranes.

We first adapted a published method for the analysis of polymer-dependent membrane deformation¹⁸. A drop of lipids is deposited and dried on a coverslip, which is then mounted on a glass slide to create a microchamber. Addition of aqueous buffers produces a reorganization of the lipids into stacks of flat membrane bilayers that were used as templates for membrane tubulation¹¹. A brain polar lipids (BPL) fraction or a synthetic lipids mixture (SLM) mimicking BPL (see Methods), both supplemented with 5% mol./mol. phosphatidyl-inositol-4,5-bisphosphate (PtdIns(4,5)P₂), were used. Injection of purified brain dynamin (0.4–1 mg ml^{−1}, no nucleotides) into the chamber induced the growth of narrow tubules, either at the edges or on the top of the most superficial layer of the lipid sheets, which could be observed by differential interference contrast (DIC) microscopy (Fig. 1a, c; Supplementary Movie S1). Tubulation required the presence of PtdIns(4,5)P₂ (compare Fig. 1a with Fig. 1b, and see Supplementary Fig. S1), consistent with the property

of dynamin to bind this phosphoinositide⁸. Tubules grew at the rate of approximately 5–7 μ m min^{−1} (Fig. 1c), and up to tens of micrometres in length (Fig. 1a), eventually forming an apparent network (Fig. 1d). Fluorescence experiments with lipids doped with rhodamine-labelled phosphatidylethanolamine (Rhod-PE) and with Alexa-488–dynamin confirmed that the linear structures visible by DIC microscopy were dynamin-coated membrane tubules (Fig. 1d). Sequential injection of Alexa-488–dynamin followed by unlabelled dynamin revealed a lack of random intermixing within the coat. Most

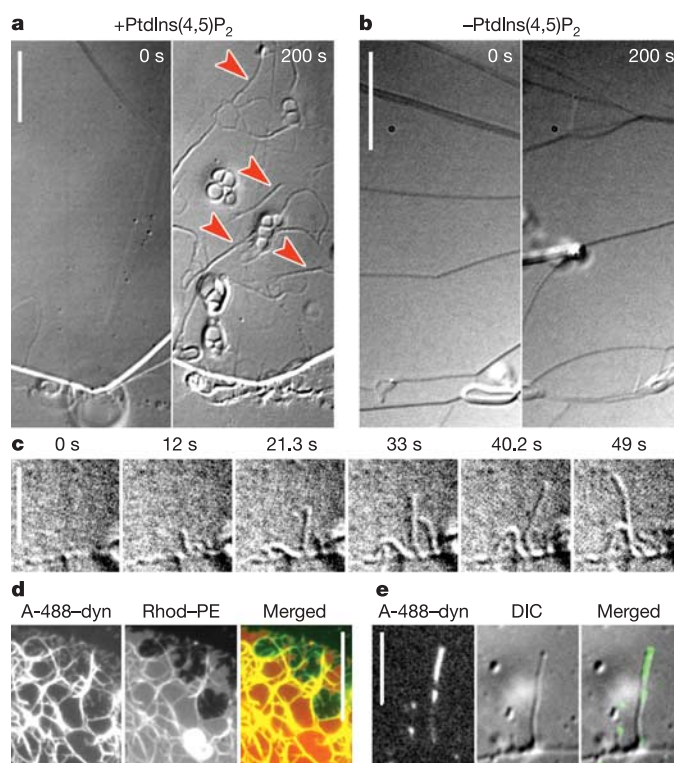


Figure 1 | Generation and growth of dynamin-coated membrane tubules.

a, Effect of dynamin (1 mg ml^{−1}) on membrane sheets composed of a synthetic lipid mixture as revealed by DIC. After dynamin addition, tubules are visible (red arrowheads). **b**, Same as in **a**, but without PtdIns(4,5)P₂. **c**, Growth of dynamin-coated tubules. **d**, Double fluorescence of tubules generated by Alexa-488–dynamin (green) on membrane sheets (SLM) containing Rhod-PE (red). **e**, Fluorescence and DIC images of a tubule generated by the sequential addition of Alexa-488–dynamin and unlabelled dynamin. Time intervals are in seconds. Scale bars: **a**, **b**, 10 μ m; **c**, 2 μ m; **e**, 5 μ m.

¹Department of Cell Biology, Howard Hughes Medical Institute, Kavli Institute for Neuroscience, and ²Department of Molecular Biochemistry and Biophysics, Boyer Center for Molecular Medicine, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

fluorescent dynamin was found at the tubule tips, suggesting that tubules grew from their base (Fig. 1e). However, fluorescent tubule segments intercalated by unlabelled segments were also seen (Fig. 1e), possibly reflecting polymerization of unlabelled dynamin from points of discontinuity within the dynamin coat.

Addition of 1 mM GTP produced dramatic changes of the network. First, most tubules became straight (within seconds), suggesting contraction and longitudinal stretching between nodes of the network. Second, they broke and collapsed, leading (within 10–20 seconds) to isolated clusters of membranes (Fig. 2a; Supplementary Movie S2). GDP caused only a light contraction of the tubules (data not shown), but not fission (Supplementary Fig. S2a). GTP- γ S did not produce either contraction or fission, but stimulated the growth of new tubules (Supplementary Fig. S2a). Experiments with fluorescent dynamin and lipids yielded the same results, demonstrating that tubule disruption by GTP is not the result of dynamin dissociation from the lipids (Supplementary Fig. S2b).

The behaviour of 93 individual tubules following GTP addition was examined. Contraction with an increase in longitudinal tension, as indicated by the acquisition of a rectilinear course (Fig. 2a, b), generally correlated with the appearance of densities along the tubule (red arrowheads in Fig. 2b; see below) and ended with a single break, followed by the retraction of the two stumps (Fig. 2b and Supplementary Fig. S2c; Supplementary Movie S3). All tubules with a free (non-anchored) end retracted without breakage (Fig. 2c and Supplementary Fig. S2d; Supplementary Movie S4). Thus, the reported GTP-dependent constricting activity of dynamin^{14,17} is insufficient to achieve fission and, at least under these *in vitro* conditions, longitudinal tension must come into play.

One could expect that, if dynamin-coated tubules were anchored at multiple sites along their length, multiple fission events would occur. Such tubules were generated by depositing giant liposomes

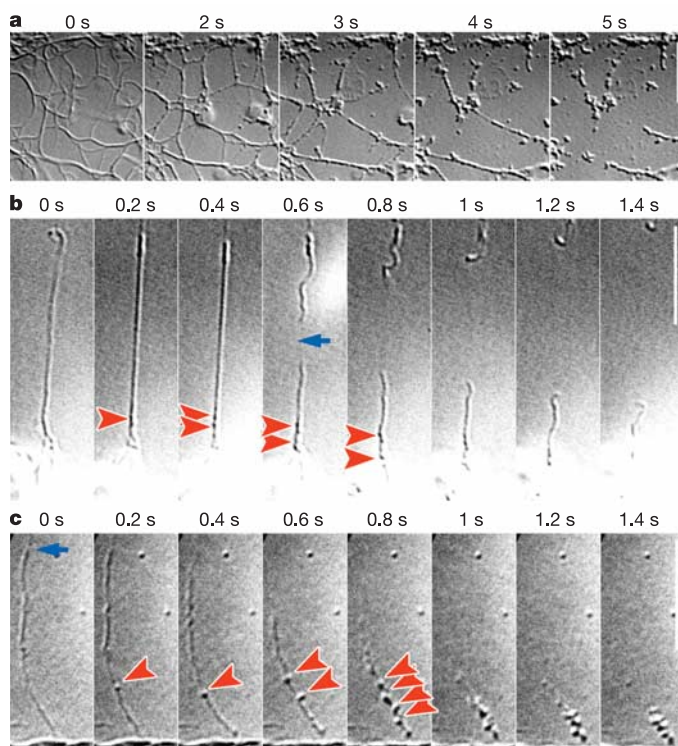


Figure 2 | Effect of guanylnucleotides on dynamin-coated tubules. **a**, 1 mM GTP induces rapid fragmentation of the tubule network, as revealed by DIC. **b**, **c**, Time-lapse sequences showing the effect of 1 mM GTP on two dynamin-coated tubules. In **b**, the tubule underwent a single break (blue arrow). In **c**, the tubule, which had a free end (blue arrow), retracted. Red arrowheads point to densities appearing along the tubules. Scale bars: 10 μ m.

doped with Rhod-PE and precoated with kinesin on a network of microtubules in the presence of ATP^{19,20} (Fig. 3a, left panel), and then adding dynamin (or Alexa-488–dynamin), in the absence of nucleotides, to the tubules pulled by kinesin. Dynamin did associate with these lipid tubules, which were bound to microtubules by multiple kinesins²⁰, and modified their appearance, but did not break them. On the basis of the overlapping fluorescence of Alexa488 and rhodamine, the tubules seemed to become thinner, with corresponding focal swellings (Fig. 3a, middle panels), in agreement with the smaller diameters of dynamin-coated tubules (approximately 20 nm, inner diameter²¹) relative to kinesin-pulled tubules (in the range of 40 nm or more¹⁹). Dynamin also bound to microtubules, consistent with its microtubule-binding properties²² (Fig. 3a, middle panels). On further addition of GTP, dynamin-coated lipid tubules (Fig. 3a, right panels, and Fig. 3b; Supplementary Movie S5 and Supplementary Fig. S3), but not naked tubules (data not shown), underwent rapid fragmentation. Most fragments remained anchored to the tubules, although some of them floated away into the medium, possibly reflecting the reversibility of kinesin–microtubule anchoring sites (Fig. 3b, red arrowheads; Supplementary Movie S5). Accordingly, such dispersion was stronger in the continued presence of ATP (data not shown). Dynamin, as expected²³, dissociated from microtubules in the presence of GTP, although it remained associated with the lipid fragments (Fig. 3a, middle and right panels). Break density along these tubules (defined as the number of breaks per micrometre

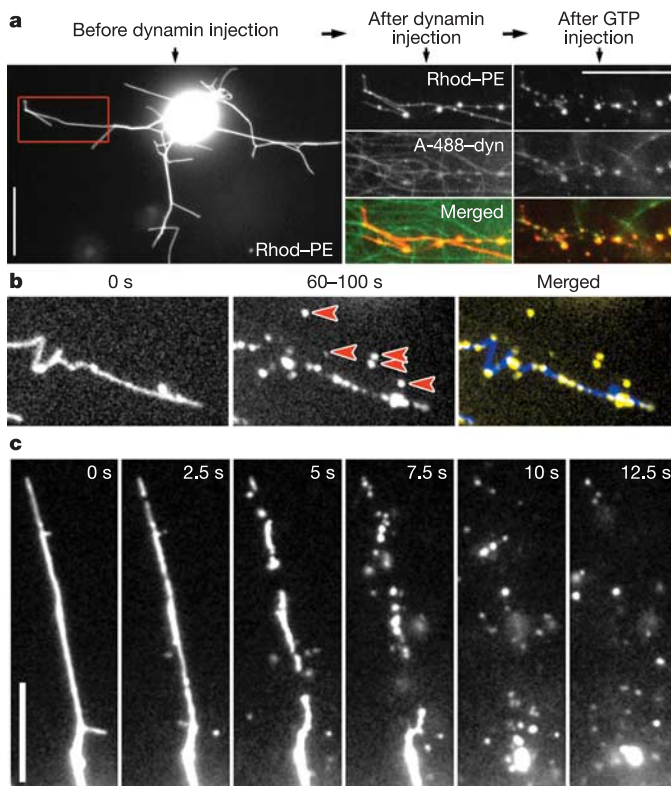


Figure 3 | Dynamin-dependent fragmentation of membrane tubules generated from giant liposomes by kinesin on a microtubule network. **a**, The lipid tubules (Rhod-PE) are shown in the left panel. A detail (red box) is shown at higher magnification in the middle and right panels after addition of Alexa-488–dynamin (1 mg ml⁻¹; middle panels) and then of 1 mM GTP (right panels), following the wash-out of ATP. Filaments positive only for Alexa 488 represents dynamin-coated microtubules. **b**, Lipid tubule (Rhod-PE) from a similar sample before (left) and after (middle; superimposition of 20 frames from 60–100 seconds) the addition of GTP. Right panel shows a merge of the two fields in pseudo-colours. **c**, Time-lapse sequence of a lipid tubule (Rhod-PE) after addition of dynamin and GTP together in the presence of ATP. Scale bars: 5 μ m.

of tubule) was ten times higher than along the tubules formed by dynamin on membrane sheets (Supplementary Fig. S3b), confirming that when tubules have multiple anchoring sites the activity of dynamin produces multiple fission events. Rapid tubule fragmentation, as well as dispersion of the fragments in the medium (more prominent in the presence of ATP; see above), was also observed when dynamin and GTP were added together (Fig. 3c and Supplementary Movie S6), indicating that dynamin-mediated fragmentation occurs very efficiently on preformed tubules that are already under tension—as they are expected to be under these conditions²⁰.

In tubules generated by dynamin on membrane sheets (Fig. 2), the tension required for fission may be generated by a GTP-hydrolysis-dependent longitudinal contraction of the dynamin coat. In fact, previous cryo-EM analysis^{17,21,24} of dynamin-coated tubules revealed that constriction also results in a reduction of the helical pitch (from 13 nm to 9 nm; ref. 17), thus producing a shortening, by approximately 30%, of the dynamin helix. This structural rearrangement of the coat is not sufficient to account for the retraction observed for free tubules in the membrane sheets assay (Fig. 2c). A potential explanation for such a dramatic shortening came from further observation of tubule dynamics on GTP addition. In several cases, coated tubules formed loops (Fig. 4a; Supplementary Movie S7), which, after GTP addition, underwent twisting to form supercoils (plectonemes; red arrowheads in Fig. 4a), suggesting an underlying twisting of the dynamin coat itself. Supercoiling could explain the retraction of free tubules and the appearance of dense spots on tubules during their contraction and/or retraction (red arrowheads in Fig. 2b, c; Supplementary Movies S3 and S4). Twisting of a helix implies partial rotations of each turn over the adjacent ones, thus resulting in complete rotations when compounded over many turns along the tubule. We investigated whether this is the case. Tubules were generated in the presence of dynamin, or biotinylated dynamin, and negatively charged beads (320 nm in diameter), or streptavidin-coated polystyrene beads (260 nm in diameter), respectively. The motility of beads attached to the tubules was then tracked by DIC microscopy (Fig. 4b, c and Supplementary Fig. S5). No motility was observed before GTP addition. GTP, after a lag phase explained, in part, by its diffusion into the microchamber, triggered both tubule tension and a marked rotation of the beads around the longitudinal axis of the tubule (Fig. 4b, c; Supplementary Movie S8). As many as 30 rotations were observed for an individual bead (see bead tracking traces in Fig. 4c), with the speed of rotation decreasing with the number of rotations until the break occurred (Supplementary Fig. S4a). The average maximal speed of bead rotation increased with the GTP concentration (Fig. 4c, d). Rotations could only be observed on 'free' tubules, such as those extending from the lipid sheets to the glass surface, as bead movement on these tubules was not hindered by adjacent structures.

To further assess the twisting activity, dynamin-coated tubules that had been reacted with GTP in suspension were deposited on grids and examined by electron microscopy. Numerous supercoils were observed (Fig. 4e), in agreement with the 'convoluted' appearance previously noted for tubules prepared in similar conditions¹⁷. The great majority of these supercoils had 'plus' configurations (Supplementary Fig. S4b), as expected from a right-handed twist²⁵. As dynamin is a right-handed helix²⁴, it will further coil, and thus constrict, during a right-handed twist (Fig. 4e). This change is consistent with previous observations made by cryo-EM: (1) snapshots of dynamin 1 in the constricted state reveal one or fewer repeating units per turn than in the dilated state²¹, and (2) both the inner diameter and the pitch of the helix decrease with GTP hydrolysis¹⁷. The increase in the pitch of the dynamin helix observed on GTP hydrolysis in one study⁸ may reflect the rigidity of the lipid nanotubes used in that particular study. The resistance to twisting opposed by the nanotubes may have induced a distortion of the dynamin helix.

Collectively, the findings reported here strongly support a mechanical action of dynamin in membrane fission. When dynamin forms an extended coat, its twisting activity also contributes to a longitudinal contraction that has been exploited in this *in vitro* study to establish the requirement of membrane tension in dynamin-mediated fission. If tubules are already under tension (Fig. 3c and Supplementary Movie S6), the constricting activity of dynamin is probably sufficient for fission. A similar situation may apply to nascent endocytic vesicles, where dynamin may form only a very short spiral (Fig. 4f). Strong biochemical and functional evidence^{11–13,26,27} has shown that the actin cytoskeleton acts together with dynamin in endocytic fission. Actin-based mechanisms may provide the required tension^{10,28}. We suggest that a cooperation of mechanisms that mediate constriction with mechanisms that create membrane tension may also occur in other types of membrane fission²⁹, although the players implicated in the generation of constriction and tension may differ in different contexts.

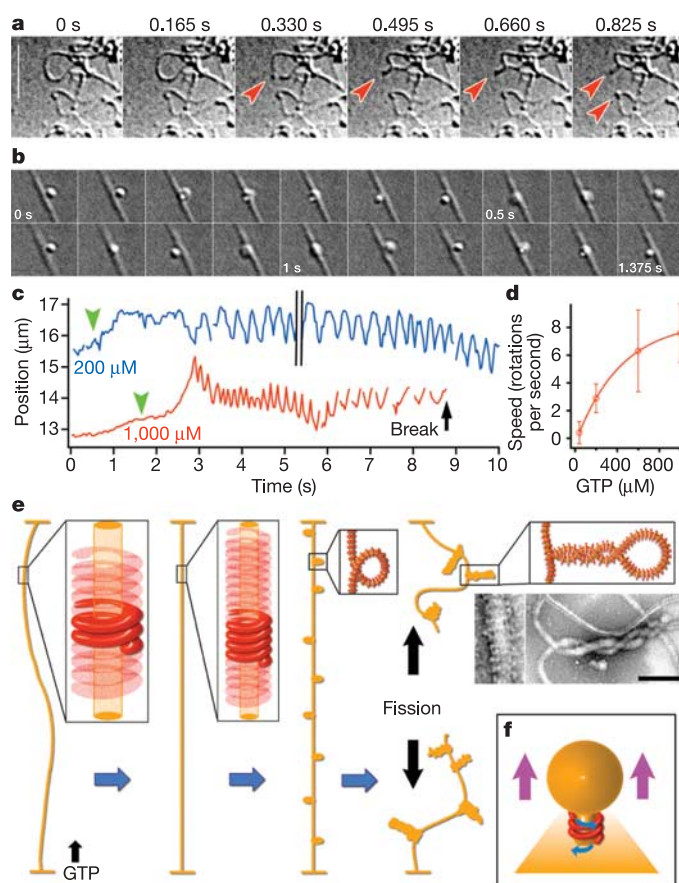


Figure 4 | Twisting activity of the dynamin coat. **a**, DIC time-lapse images showing loops of dynamin-coated tubules undergoing supercoiling (red arrowheads) on the addition of 1 mM GTP. **b**, Effect of 200 μ M GTP on a tubule generated by biotinylated dynamin in the presence of streptavidin-coated polystyrene beads. **c**, Tracking of two beads at different GTP concentrations. The y axis depicts the displacement perpendicular to the axis of the tubules. Green arrowheads indicate the time of GTP addition. **d**, Average ($\pm$ s.d.) maximal angular speed of the beads as a function of the GTP concentration. **e**, Proposed effect of GTP on dynamin-coated tubules. The dynamin helix undergoes an increase in torsion leading to straightening of the tubule and supercoiling until a break occurs allowing more supercoiling. An electron microscopy image (and a detail showing dynamin rings) of a supercoiled dynamin-coated tubule following incubation with 500 μ M GTP is also shown. **f**, Proposed cooperation in the fission of an endocytic vesicle of the twisting action of dynamin (blue arrows), which produces constriction, and of factors that induce tension (purple arrows). Scale bars: **a**, 10 μ m; **b**, 1 μ m; **e**, 200 nm.

METHODS

Materials. All lipids were purchased from Avanti Polar Lipids. Two lipid preparations were used: 95% brain polar lipids (BPL) plus 5% mol./mol. phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂), and a synthetic lipid mixture (SLM) composed of 81% mol./mol. phospholipids (30% brain phosphatidylethanolamine (PE), 5% liver phosphoinositides (PI), 30% palmitoyl-oleyl phosphatidylserine (POPS), and 35% palmitoyl-oleyl phosphatidylcholine (POPC)), 14% mol./mol. cholesterol, and 5% mol./mol. PtdIns(4,5)P₂. In control experiments, PtdIns(4,5)P₂ was omitted.

Dynamin purification and labelling. Dynamin was purified from rat brains using the GST-tagged SH3 domain of rat amphiphysin 2 as an affinity ligand as described^{8,16}. Conjugation of dynamin to Alexa 488 or biotin was carried out by standard procedures. Unlabelled and labelled dynamins were dialysed against storage buffer (20 mM HEPES, pH 7.4, 100 mM NaCl, 50% (vol./vol.) glycerol), aliquoted and stored at -80°C.

Formation of membrane sheets. A prewashed glass coverslip stored in ethanol was dried under a N₂ flux, and 1 µl droplets of lipid solution (10 mg ml⁻¹ in pure chloroform) were deposited and allowed to dry on the coverslip. Lipid-coated coverslips were further dried under vacuum for at least one hour. A small chamber was built by placing the coverslip onto a glass slide, with the lipids facing the glass slide, using double-sided tape as spacers (see Fig. 2 in ref. 11). The lipids were fully rehydrated by filling the chamber by capillary action with 15–20 µl of buffer (0.1 mg ml⁻¹ casein in 20 mM HEPES, pH 7.4, 100 mM NaCl, 1 mM MgCl₂). Five microlitres of a dynamin solution was then applied and the deformation of membrane sheets was recorded by DIC microscopy at a high resolution (1,300 × 1,024 pixels) video rate (30 frames per second). Five microlitres of nucleotides containing buffer were added after formation of the tubes. All incubations were carried out at room temperature (approximately 22°C).

Kinesin-induced lipid tubulation. Microtubule-coated chambers between glass slides were prepared as described²⁰. Giant liposomes were generated using a modified protocol from ref. 30. Tubules were generated as follows: 3 µl of giant liposomes were mixed with 2 µl of 0.1 mg ml⁻¹ streptavidin in casein buffer and incubated for 3 min. Two microlitres of biotinylated recombinant kinesin (the plasmid was a gift from P. Bassereau) was then added, incubated for 5 min and mixed with 7 µl of motility buffer (2 mM ATP, an oxygen scavenger enzymatic system (see Supplementary Methods for details), 10 µM taxol in casein buffer)²⁰. This mix was then added to the microtubule-coated chamber. After a 15–20 min incubation to allow tubule growth, subsequent injections (buffers, dynamin, nucleotides) were performed. All incubations were carried out at room temperature (approximately 22°C).

Received 30 November 2005; accepted 13 March 2006.

Published online 30 April 2006.

- Koenig, J. H. & Ikeda, K. Disappearance and reformation of synaptic vesicle membrane upon transmitter release observed under reversible blockage of membrane retrieval. *J. Neurosci.* **9**, 3844–3860 (1989).
- Sever, S., Damke, H. & Schmid, S. L. Garrotes, springs, ratchets, and whips: putting dynamin models to the test. *Traffic* **1**, 385–392 (2000).
- Thompson, H. M. & McNiven, M. A. Dynamin: switch or pinchase? *Curr. Biol.* **11**, R850 (2001).
- Takei, K., McPherson, P. S., Schmid, S. L. & De Camilli, P. Tubular membrane invaginations coated by dynamin rings are induced by GTP-γS in nerve terminals. *Nature* **374**, 186–190 (1995).
- Hinshaw, J. E. & Schmid, S. L. Dynamin self-assembles into rings suggesting a mechanism for coated vesicle budding. *Nature* **374**, 190–192 (1995).
- Praefcke, G. J. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature Rev. Mol. Cell Biol.* **5**, 133–147 (2004).
- Sever, S., Muhlberg, A. B. & Schmid, S. L. Impairment of dynamin's GAP domain stimulates receptor-mediated endocytosis. *Nature* **398**, 481–486 (1999).
- Stowell, M. H., Marks, B., Wigge, P. & McMahon, H. T. Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring. *Nature Cell Biol.* **1**, 27–32 (1999).
- Slepnev, V. I. & De Camilli, P. Accessory factors in clathrin-dependent synaptic vesicle endocytosis. *Nature Rev. Neurosci.* **1**, 161–172 (2000).
- Merrifield, C. J., Perrais, D. & Zenisek, D. Coupling between clathrin-coated-pit invagination, cortactin recruitment, and membrane scission observed in live cells. *Cell* **121**, 593–606 (2005).
- Itoh, T. *et al.* Dynamin and the actin cytoskeleton cooperatively regulate plasma membrane invagination by BAR and F-BAR proteins. *Dev. Cell* **9**, 791–804 (2005).
- Qualmann, B., Kessels, M. M. & Kelly, R. B. Molecular links between endocytosis and the actin cytoskeleton. *J. Cell Biol.* **150**, F111–F116 (2000).
- Cao, H. *et al.* Cortactin is a component of clathrin-coated pits and participates in receptor-mediated endocytosis. *Mol. Cell Biol.* **23**, 2162–2170 (2003).
- Sweitzer, S. M. & Hinshaw, J. E. Dynamin undergoes a GTP-dependent conformational change causing vesiculation. *Cell* **93**, 1021–1029 (1998).
- Takei, K., Slepnev, V. I., Haucke, V. & De Camilli, P. Functional partnership between amphiphysin and dynamin in clathrin-mediated endocytosis. *Nature Cell Biol.* **1**, 33–39 (1999).
- Farsad, K. *et al.* Generation of high curvature membranes mediated by direct endophilin bilayer interactions. *J. Cell Biol.* **155**, 193–200 (2001).
- Danino, D., Moon, K. H. & Hinshaw, J. E. Rapid constriction of lipid bilayers by the mechanochemical enzyme dynamin. *J. Struct. Biol.* **147**, 259–267 (2004).
- Tsafir, I., Caspi, Y., Guedeau-Boudeville, M. A., Arzi, T. & Stavans, J. Budding and tubulation in highly oblate vesicles by anchored amphiphilic molecules. *Phys. Rev. Lett.* **91**, 138102 (2003).
- Roux, A. *et al.* A minimal system allowing tubulation with molecular motors pulling on giant liposomes. *Proc. Natl Acad. Sci. USA* **99**, 5394–5399 (2002).
- Leduc, C. *et al.* Cooperative extraction of membrane nanotubes by molecular motors. *Proc. Natl Acad. Sci. USA* **101**, 17096–17101 (2004).
- Chen, Y. J., Zhang, P., Egelman, E. H. & Hinshaw, J. E. The stalk region of dynamin drives the constriction of dynamin tubes. *Nature Struct. Mol. Biol.* **11**, 574–575 (2004).
- Shpetner, H. S. & Vallee, R. B. Identification of dynamin, a novel mechanochemical enzyme that mediates interactions between microtubules. *Cell* **59**, 421–432 (1989).
- Obar, R. A., Collins, C. A., Hammarback, J. A., Shpetner, H. S. & Vallee, R. B. Molecular cloning of the microtubule-associated mechanochemical enzyme dynamin reveals homology with a new family of GTP-binding proteins. *Nature* **347**, 256–261 (1990).
- Zhang, P. & Hinshaw, J. E. Three-dimensional reconstruction of dynamin in the constricted state. *Nature Cell Biol.* **3**, 922–926 (2001).
- Charvin, G., Bensimon, D. & Croquette, V. Single-molecule study of DNA unlinking by eukaryotic and prokaryotic type-II topoisomerases. *Proc. Natl Acad. Sci. USA* **100**, 9820–9825 (2003).
- Orth, J. D. & McNiven, M. A. Dynamin at the actin-membrane interface. *Curr. Opin. Cell Biol.* **15**, 31–39 (2003).
- Schafer, D. A. Regulating actin dynamics at membranes: a focus on dynamin. *Traffic* **5**, 463–469 (2004).
- Kaksonen, M., Toret, C. P. & Drubin, D. G. A modular design for the clathrin- and actin-mediated endocytosis machinery. *Cell* **123**, 305–320 (2005).
- Roux, A. *et al.* Role of curvature and phase transition in lipid sorting and fission of membrane tubules. *EMBO J.* **24**, 1537–1545 (2005).
- Moscho, A., Orwar, O., Chiu, D. T., Modi, B. P. & Zare, R. N. Rapid preparation of giant unilamellar vesicles. *Proc. Natl Acad. Sci. USA* **93**, 11443–11447 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Unger, G. Di Paolo, M. Solimena and K. Erdmann for discussion and critical reading of this manuscript. We thank F. Wilson for technical support. This work was supported by National Institutes of Health grants and a G. Harold and Leila Y. Mathers Charitable Foundation grant to P.D.C. A.R. was supported by the European Molecular Biology Organization (EMBO) Long-Term Postdoctoral Fellowship programme and the Cross-Disciplinary Fellowship programme of the Human Frontier Science Program (HFSP).

Author Contributions A.R. and P.D.C. conceived the project, designed the experiments and evaluated the results. A.R. performed the experiments alone, with the exception of the giant-liposome assay (K.U. and A.R.) and electron microscopy (A.F., K.U. and A.R.). A.R. and P.D.C. wrote the paper.

Author Information Reprints and permissions information is available at ngp.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.D.C. (piero.decamilli@yale.edu).

LETTERS

Rec8 phosphorylation and recombination promote the step-wise loss of cohesins in meiosis

Gloria A. Brar¹, Brendan M. Kiburz¹, Yi Zhang², Ji-Eun Kim², Forest White² & Angelika Amon¹

During meiosis, cohesins—protein complexes that hold sister chromatids together—are lost from chromosomes in a step-wise manner¹. Loss of cohesins from chromosome arms is necessary for homologous chromosomes to segregate during meiosis I. Retention of cohesins around centromeres until meiosis II is required for the accurate segregation of sister chromatids. Here we show that phosphorylation of the cohesin subunit Rec8 contributes to step-wise cohesin removal. Our data further implicate two other key regulators of meiotic chromosome segregation, the cohesin protector Sgo1 and meiotic recombination in bringing about the step-wise loss of cohesins and thus the establishment of the meiotic chromosome segregation pattern. Understanding the interplay between these processes should provide insight into the events underlying meiotic chromosome mis-segregation, the leading cause of miscarriages and mental retardation in humans.

During the meiotic cell cycle, DNA replication is followed by two rounds of chromosome segregation, in which homologues segregate during the first division and sister chromatids are partitioned in the second. Loss of cohesins from chromosome arms allows the segregation of homologous chromosomes during meiosis I by causing the resolution of meiotic recombination events, which hold homologous chromosomes together before anaphase I (ref. 2). Maintenance of cohesins around centromeres beyond anaphase I and cohesin removal during meiosis II are essential for accurate segregation of sister chromatids. Members of the MEI-S332/Sgo1 (Shugoshin) protein family protect cohesins around centromeres from removal during meiosis I by recruiting protein phosphatase 2A (PP2A) to centromeric regions^{3–10}.

Cohesins are removed from chromosomes by a protease known as separase (Esp1 in yeast). After the ubiquitin-dependent destruction of its inhibitor securin (Pds1 in yeast) mediated by the anaphase promoting complex/cyclosome (APC/C), separase cleaves a subunit of the cohesin complex, thereby triggering anaphase¹¹. The Polo kinase Cdc5 contributes to cohesin removal by promoting cleavage-independent cohesin dissociation during meiotic prophase¹² and cohesin cleavage by separase^{13–16}. Phosphorylation of the cohesin subunit and separase target Rec8 is furthermore decreased in cells lacking Cdc5 (ref. 14), raising the possibility that Rec8 phosphorylation is important for the protein's cleavage.

To determine the importance of Rec8 phosphorylation in cohesin cleavage we mapped the phosphorylation sites of Rec8 isolated from cells arrested in metaphase I either by depletion of the APC/C activator Cdc20 (ref. 14), or by expression of a non-degradable version of Pds1 from the meiosis-specific *DMC1* promoter (Supplementary Fig. 1). In both arrests, Rec8 is highly phosphorylated (data not shown). We also isolated Rec8 from cells depleted for Cdc5 to identify Cdc5-dependent phosphorylation events. The overall coverage of Rec8 was 85% (Supplementary Fig. 2) leading to the identification of 24 phosphorylation sites (Supplementary Table 1;

Supplementary Information). Eleven of these sites were phosphorylated in Cdc20-depleted cells or Pds1 Δ -expressing cells but not in Cdc5-depleted cells, suggesting that these sites are phosphorylated by Cdc5 *in vivo*. This analysis defined the motif S/E/D-X_{0–2}-N(Q)-X_{0–2}-Sp(Tp)-X₃- π (where π represents a polar amino acid, and p indicates a phosphorylated amino acid residue) as the consensus sequence for Cdc5-dependent phosphorylation sites (Supplementary Table 2). Additional features of the region surrounding these sites are discussed in the Supplementary Information.

We mutated the phosphorylated sites within Rec8 to amino acids that can no longer be phosphorylated. Mutation of individual phosphorylation sites to alanine did not affect sporulation efficiency (data not shown). Mutation of multiple phosphorylation sites to alanine led to a delay in prophase I (Supplementary Fig. 3a), which is indicative of a partial loss of Rec8 function¹⁷. This result suggests that phosphorylation of Rec8 may be important for the protein's prophase functions.

Next, we examined the consequences of mutating the 11 residues whose phosphorylation was Cdc5-dependent (*rec8-psa*) to alanine. A defect in cohesin removal is expected to interfere with anaphase I entry¹⁸ but cells expressing this allele did not exhibit a defect at this transition (Supplementary Fig. 3b), suggesting either that Cdc5-dependent phosphorylation of Rec8 was not important for Rec8 removal or that our mass-spectrometry analysis did not identify all Cdc5-dependent phosphorylation sites. The latter scenario was more likely, given that the coverage in the Cdc5-depletion arrest was only 65%. We therefore—in addition to the known Cdc5-dependent sites—mutated putative Cdc5-dependent phosphorylation sites that were found to be phosphorylated in the *pCLB2-CDC20* and/or *pDMC1-PDS1 Δ* arrests but that were not covered in the Cdc5-depletion arrest, as well as three Cdc5-independent sites, to alanine (*rec8-17A*; Supplementary Table 1). Cells expressing this *rec8* mutant exhibited a 1-h prophase delay and a metaphase I delay that was not as great as that observed in cells expressing a non-cleavable version of Rec8 (Fig. 1a, Supplementary Fig. 4a). Entry into anaphase II was only slightly—if at all—delayed (Supplementary Fig. 4b), suggesting that Rec8 phosphorylation is less important for this cell cycle transition. *rec8* mutants in which all phosphorylated serines and threonines, except two recently identified sites (T291, S292), were mutated to alanine (*rec8-21A*) and mutants that had additional putative Cdc5 phosphorylation sites mutated to alanine (*rec8-24A*, *rec8-29A*; Supplementary Table 1) also appeared to be delayed in metaphase I, although the extent of the delay was difficult to assess owing to the severe prophase I delay exhibited by the mutants (Supplementary Fig. 3c). We conclude that mutating Rec8's phosphorylation sites leads to impairment in Rec8's prophase function and interferes with anaphase I entry.

The *rec8-17A* mutant was analysed in more detail because the

¹Center for Cancer Research, Howard Hughes Medical Institute, Massachusetts Institute of Technology, E17-233 40 Ames Street, Cambridge, Massachusetts 02139, USA.

²Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Building 56-787a, Cambridge, Massachusetts 02139-4307, USA.

metaphase I delay was the least obscured by the prophase delay in this mutant and because the mutant was likely to have most of its Cdc5-dependent phosphorylation sites mutated to alanine. Analysis of Pds1 by indirect *in situ* immunofluorescence revealed that *rec8-17A* cultures contained a significant fraction of metaphase I cells lacking Pds1 (Fig. 1c, Supplementary Fig. 5). We conclude that the metaphase I delay observed in the *rec8-17A* mutant is at least in part due to events occurring after the degradation of Pds1.

Next we determined whether the metaphase I delay observed in *rec8-17A* mutants was due to a Rec8 cleavage defect. In wild-type cells, the carboxy-terminal Rec8 cleavage product accumulated 4 h after induction of sporulation (Fig. 1b). The Rec8-17A protein assembled onto chromosomes normally (Fig. 1d, e) but cleavage did not occur until 7 h (Fig. 1a, b). This delay was only due in part to

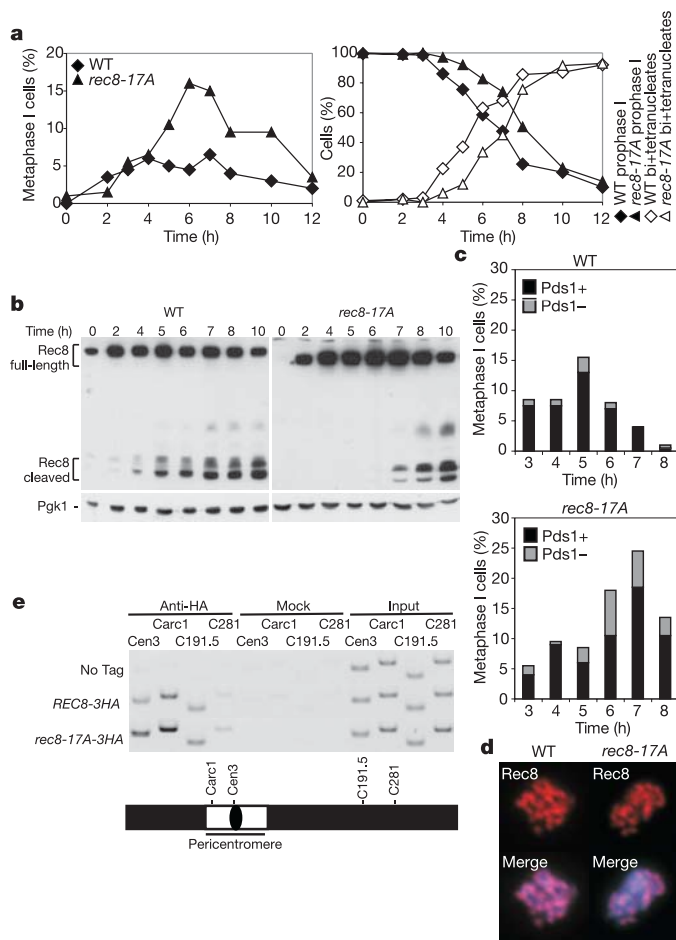


Figure 1 | Rec8 cleavage is delayed in *rec8-17A* cells. **a**, **b**, Wild-type (WT) (A14655, diamonds) and *rec8-17A* mutant (A14746, triangles) cells lacking the ubiquitin ligase Ubr1 to facilitate detection of the Rec8 cleavage product were induced to sporulate to determine the percentage of metaphase I cells (**a**, left panel), of prophase I (**a**, right panel, solid symbols) and of the sum of bi- and tetra-nucleate cells (**a**, right panel, open symbols). Rec8-3HA and Pgs1 (loading control) were analysed by western blotting (**b**). **c**, WT (A14923) and *rec8-17A* mutant (A14861) cells both carrying a *PDS1-18MYC* fusion were induced to sporulate and Pds1 staining was determined in all metaphase I cells. **d**, The localization of Rec8 is shown on chromosome spreads of WT cells and *rec8-17A* mutants. Rec8 is shown in red, DNA in blue in the merge. **e**, WT *REC8-3HA* (A1972) and *rec8-17A-3HA* (A13559) and a control strain (A4962) were induced to sporulate and processed for chromatin immunoprecipitation after 4 h. Polymerase chain reaction (PCR) analysis of immunoprecipitated samples (anti-haemagglutinin; anti-HA), mock-treated samples, and input DNA (1:250) are shown along with a diagram indicating locations of chromosome III primer sets. C281 is a cohesin-poor location.

the prophase I delay, which was 90 min (Fig. 1a). Our results indicate that cleavage of the Rec8-17A mutant protein is delayed not only owing to prophase I defects but also owing to defects occurring after the degradation of Pds1. We conclude that phosphorylation of Rec8 is important for its timely cleavage during meiosis I.

We also examined the effects of eliminating meiotic recombination on *rec8-17A* mutants. Surprisingly, deletion of *SPO11*, a gene required to form the recombination-initiating double-strand breaks^{19,20} or expression of a catalytic dead version of Spo11 (*spo11-Y135F*)¹⁹ abolished the delay in Rec8 cleavage and cell cycle progression of *rec8-17A* or *rec8-29A* mutants (Fig. 2a, Supplementary Fig. 6, 7). Our results suggest that in the absence of recombination Rec8 phosphorylation is not as important for cohesin removal as it is when recombination occurs.

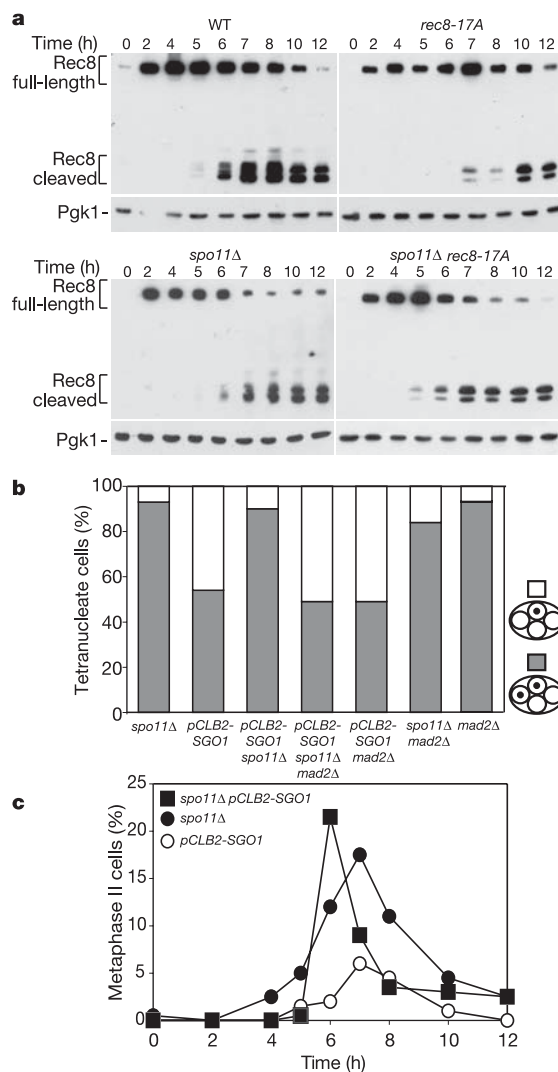


Figure 2 | Elimination of recombination abolishes the Rec8 cleavage delay in *rec8-17A* owing to retention of arm cohesion past meiosis I. **a**, WT (A14655), *spo11Δ* (A14755), *rec8-17A* (A14746) and *spo11Δ rec8-17A* (A14847) cells were induced to sporulate and Rec8-3HA was analysed by western blotting. **b**, *spo11Δ* (A9498), *spo11Δ pCLB2-SGO1* (A14938), *pCLB2-SGO1* (A11251), *spo11Δ pCLB2-SGO1 mad2Δ* (A15345), *spo11Δ mad2Δ* (A15490), *mad2Δ* (A15494) and *pCLB2-SGO1 mad2Δ* (A15344) cells carrying CEN5-GFP dots were induced to sporulate to determine GFP dot segregation in tetrads ($n = 100$). **c**, *spo11Δ* (A9498, closed circles), *spo11Δ pCLB2-SGO1* (A14938, squares) and *pCLB2-SGO1* (A11251, open circles) cells were induced to sporulate to determine the percentage of metaphase II cells.

Why does elimination of recombination suppress the Rec8 cleavage defect of *rec8-17A* mutants? In *spo11Δ*, *spo11Δ rec8-17A* and *spo11Δ rec8-29A* mutants, loss of cohesins from chromosome arms and from centromeric regions occurs almost simultaneously, as is shown by the absence of binucleate cells with cohesins concentrated around

centromeres in chromosome spreads (Supplementary Fig. 8a–d). This raises the possibility that in this mutant the bulk of cohesin removal occurs during meiosis II, when phosphorylation appears less important for Rec8 cleavage (Supplementary Fig. 4b). To test this hypothesis we examined the effects of deleting *SPO11* in Sgo1-depleted cells.

In Sgo1-depleted cells, the second meiotic division is random, owing to the absence of cohesion between sister chromatids. This phenotype can be observed when cells carry a tandem array of Tet operator sequences near the centromere on one of the two homologues and also express a Tet repressor–green fluorescent protein (TetR–GFP) fusion that binds to these repeats (*CEN5 GFP dots*²¹). When meiosis II segregation is random, 50% of tetrads contain GFP dots in only one of the four spores and 50% of tetrads contain a GFP signal in two of the four spores (Fig. 2b, Supplementary Fig. 8d). Remarkably, 90% of *spo11Δ pCLB2-SGO1* cells segregate sister chromatids correctly (a GFP signal in two of the four spores; Fig. 2b). Furthermore,

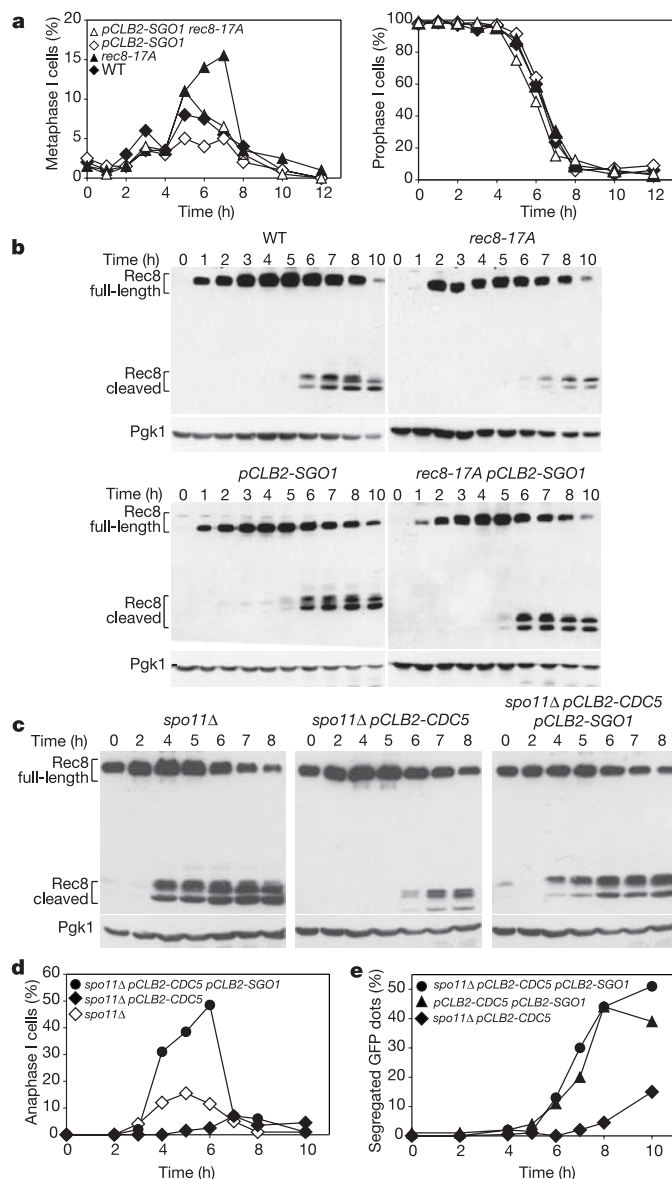


Figure 3 | Depletion of Sgo1 partially alleviates the need for Rec8 phosphorylation and Cdc5 in Rec8 cleavage and anaphase I entry. **a**, **b**, WT (A15086, closed diamonds), *rec8-17A* (A14750, closed triangles), *pCLB2-SGO1* (A15085, open diamonds) and *pCLB2-SGO1 rec8-17A* (A15084, open triangles) were induced to sporulate to determine the percentage of metaphase I cells (**a**, left panel), prophase I cells (**a**, right panel) and Rec8-HA protein by western blot analysis (**b**). **c**, **d**, *spo11Δ* (A15022, open diamonds), *spo11Δ pCLB2-CDC5* (A15025, closed diamonds), and *spo11Δ pCLB2-CDC5 pCLB2-SGO1* (A15000, closed circles) cells were induced to sporulate to determine the percentage of anaphase I cells (**d**) and Rec8 protein levels (**c**). **e**, *spo11Δ pCLB2-CDC5* (A14657, closed diamonds), *pCLB2-CDC5 pCLB2-SGO1* (A14870, closed triangles), and *spo11Δ pCLB2-CDC5 pCLB2-SGO1* (A14776, closed circles) cells all carrying *CEN5-GFP* dots were induced to sporulate to determine the percentage of cells with GFP dots separated by at least 2 μ m ($n = 200$ per time point). Note that in metaphase I-arrested *spo11Δ pCLB2-CDC5* mutants, two juxtaposed GFP dots are visible because sister kinetochores attach to opposite poles rather than the same pole in meiosis I and the tension exerted by the spindle leads to separation of *CEN5 GFP* dots^{14,15}.

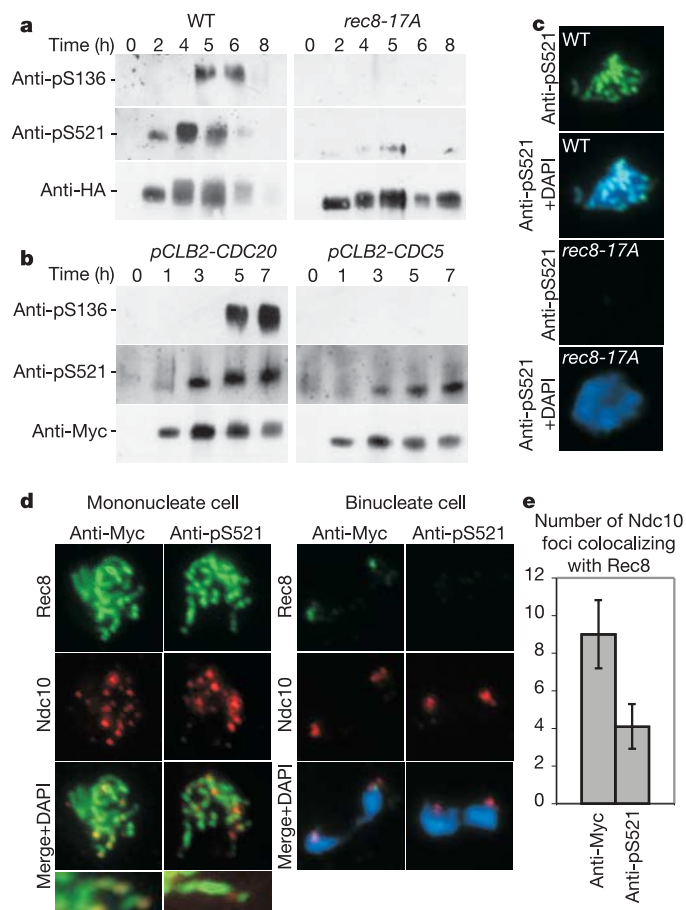


Figure 4 | Serine 521 phosphorylation is reduced around centromeres during meiosis I. **a**, WT (A1972) and *rec8-17A* mutant (A13559) cells were induced to sporulate and Rec8-HA immunoprecipitates were probed with either anti-HA or anti-phospho S136 (anti-pS136) or anti-phospho S521 (anti-pS521) antibodies. **b**, Rec8-Myc was immunoprecipitated from *pCLB2-CDC20* (A5441) and *pCLB2-CDC5* (A9858) cells and probed with either anti-Myc or anti-pS136 or anti-pS521 antibodies. **c**, Anti-pS521 staining was analysed on chromosome spreads in WT (A14655) and *rec8-17A* mutant (A14746) cells 4 h after sporulation induction. Anti-pS521 staining is shown in green and DNA in blue. **d**, **e**, WT cells carrying a *REC8-MYC* fusion and a *NDC10-HA* fusion (A3640) were spread and the distribution of Rec8 was determined either using anti-Myc or anti-pS521 antibodies. Examples of prophase and binucleate cells are shown. Rec8 is shown in green, Ndc10 in red and DNA in blue. **e**, The number of Ndc10 foci overlapping with the anti-Myc and anti-pS521 staining ($n = 12$). Error bars represent standard deviation.

deletion of *SPO11* restored metaphase II to Sgo1-depleted cells (Fig. 2c, Supplementary Fig. 8e). Similar results were obtained with other recombination mutants, such as *spo11-YF*, *rad50S* (double-strand break resection defective²²) or *dmc1Δ* (strand invasion defective²³) mutants in which recombination-induced linkages between homologues are abolished (Supplementary Fig. 9). This observation, together with the finding that chromosome segregation was again random in *spo11ΔpCLB2-SGO1mad2Δ* triple mutants (Fig. 2b), provided insight into why cohesin removal did not occur during meiosis I in the absence of recombination: in the absence of linkages between homologues, chromosomes fail to attach properly to the meiosis I spindle. This leads to the activation of the spindle assembly checkpoint, which in turn prevents the removal of cohesins from chromosomes. Cells nevertheless undergo anaphase I and progress into meiosis II because chromosomes lack the necessary linkages to prevent meiosis I spindle elongation^{24,25}. This results in metaphase II chromosomes with cohesins on chromosome arms. These observations, together with the finding that Rec8 phosphorylation is not important for Rec8 cleavage during meiosis II, explain why elimination of recombination abolishes the Rec8 cleavage delay in the *rec8-17A* mutant and demonstrate a role for recombination in establishing the step-wise loss of cohesins from chromosomes.

Sgo1 was shown to protect cohesins by recruiting PP2A to chromosomes^{9,10}. If Sgo1 solely functioned to prevent centromeric cohesin removal by preventing the phosphorylation of Rec8, inactivation of *SGO1* should not affect the phenotype exhibited by Rec8-17A-expressing cells. Surprisingly, depletion of Sgo1 in Rec8-17A-expressing cells led to Rec8 cleavage, almost to the extent seen in wild-type cells, and an elimination of the metaphase I delay (Fig. 3a, b, Supplementary Fig. 10a) implicating Sgo1 in functions other than preventing Rec8 phosphorylation. An alternative explanation for this observation was that our mass-spectrometry analysis missed key phosphorylation sites, phosphorylation of which would allow for efficient cleavage of the Rec8-17A mutant protein in the absence of Sgo1. To test this possibility we examined Rec8 cleavage in cells depleted for Cdc5. Depletion of Sgo1 allowed Rec8 cleavage, and loss of Rec8 from chromosomes occurred (Fig. 3c; Supplementary Fig. 11). Chromosome segregation, as judged by the separation of *CEN5* GFP dots, and spindle elongation also took place (Fig. 3d, e, Supplementary Fig. 10b–d). It is possible that in the absence of Sgo1, low levels of Cdc5 and other protein kinases are capable of bringing about cohesin removal. We consider this possibility unlikely because both sister chromatid separation and spindle elongation occur with remarkable efficiency. Instead we suggest that Sgo1 affects cohesin cleavage by means in addition to preventing Rec8 phosphorylation.

To determine whether Rec8 phosphorylation contributes to establishing the step-wise nature of cohesin removal we raised two antibodies, one that recognizes phospho-serine 136 and one that recognizes phospho-serine 521 (Fig. 4a, Supplementary Fig. 12a). As predicted by the mass-spectrometry analysis, phosphorylation of S136 is Cdc5-dependent and phosphorylation of S521 is Cdc5-independent (Fig. 4a, b; Supplementary Fig. 12b). The anti-phospho-S521 antibody recognized phospho-S521 on chromosome spreads (Fig. 4c) and revealed that Rec8 phosphorylation on S521 mirrored the differential loss of arm and centromeric cohesins during meiosis I. Rec8 visualized using an antibody against a C-terminal tag (Rec8-Myc) appeared continuous and was found in long stretches on chromosome spreads, presumably representing chromosome axes. In contrast, the anti-phospho-S521 signal appeared fragmented (Fig. 4d) and frequently did not overlap with the kinetochore marker Ndc10 (Fig. 4e). Furthermore, the anti-phospho-S521 signal was absent from chromosome spreads of binucleate cells when only centromeric cohesins are left on chromosomes (Fig. 4d). S521 phosphorylation was not affected by depletion of Sgo1 or deletion of the centromere-associated kinase *BUB1* (data not shown). Our results show that S521 phosphorylation is reduced or perhaps even excluded from centromeric regions, but present on chromosome

arms. However, Sgo1 and Bub1 do not appear to regulate the phosphorylation state of S521 either because they only regulate the phosphorylation state of a subset of Rec8 phosphorylation sites or because they affect cohesins at centromeric regions through means other than preventing Rec8 phosphorylation. We do not know whether Rec8 is phosphorylated before its removal in metaphase II. We have not detected an anti-phospho-S521 signal in any binucleate cells. This result suggests that Rec8 phosphorylation on S521 may not be a prerequisite for Rec8 removal during meiosis II, which would be consistent with the observation that the *rec8-17A* mutant does not exhibit a delay in metaphase II.

Our studies not only produced an *in vivo*-derived consensus sequence for targets of Polo kinases but also provided insights into how cohesin removal is regulated in meiosis. Our results suggest that it is overall phosphorylation rather than phosphorylation of a specific site that is important for Rec8 cleavage. We also observed that the delay in Rec8 cleavage in the *rec8-17A* mutant was significantly shorter than that for cells lacking Cdc5, probably because cells depleted for Cdc5 exhibit additional defects^{14,15}, or because additional Cdc5 phosphorylation sites may exist in Rec8. Our results also revealed a role for recombination in establishing the step-wise loss of cohesion. Recombination establishes linkages between homologues, which are essential for silencing of the spindle checkpoint and thus the timely removal of cohesins from chromosome arms. Thus recombination not only ensures the correct attachment of bivalents to the meiosis I spindle but also, together with Rec8 phosphorylation and Sgo1, establishes the step-wise loss of cohesion, another key aspect of meiotic chromosome segregation.

METHODS

Strains and plasmids used in this study are described in Supplementary Table 3 and Supplementary Methods, respectively. Immunoblots were performed as described in ref. 26. Chromatin immunoprecipitation (ChIP) was performed according to ref. 8. Chromosomes were spread according to ref. 27 and indirect *in situ* immunofluorescence was performed according to ref. 28.

Received 22 December 2005; accepted 10 April 2006.

Published online 3 May 2006.

- Marston, A. L. & Amon, A. Meiosis: cell-cycle controls shuffle and deal. *Nature Rev. Mol. Cell Biol.* **5**, 983–997 (2004).
- Buonomo, S. B. *et al.* Division of the nucleolus and its release of Cdc14 during anaphase of meiosis I depends on Separase, Spo12, and Slk19. *Dev. Cell* **4**, 727–739 (2003).
- Tang, T. T., Bickel, S. E., Young, L. M. & Orr-Weaver, T. L. Maintenance of sister-chromatid cohesion at the centromere by the *Drosophila* MEI-5332 protein. *Genes Dev.* **12**, 3843–3856 (1998).
- Kerrebrock, A. W., Moore, D. P., Wu, J. S. & Orr-Weaver, T. L. MEI-5332, a *Drosophila* protein required for sister-chromatid cohesion, can localize to meiotic centromere regions. *Cell* **83**, 247–256 (1995).
- Katis, V. L., Galova, M., Rabitsch, K. P., Gregan, J. & Nasmyth, K. Maintenance of cohesin at centromeres after meiosis I in budding yeast requires a kinetochore-associated protein related to MEI-5332. *Curr. Biol.* **14**, 560–572 (2004).
- Kitajima, T. S., Kawashima, S. A. & Watanabe, Y. The conserved kinetochore protein Shugoshin protects centromeric cohesion during meiosis. *Nature* **427**, 510–517 (2004).
- Marston, A. L., Tham, W. H., Shah, H. & Amon, A. A genome-wide screen identifies genes required for centromeric cohesion. *Science* **303**, 1367–1370 (2004).
- Kiburz, B. M. *et al.* The core centromere and Sgo1 establish a 50-kb cohesin-protected domain around centromeres during meiosis I. *Genes Dev.* **19**, 3017–3030 (2005).
- Kitajima, T. S. *et al.* Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **441**, 46–52 (2006).
- Riedel, C. G. *et al.* Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature* **441**, 53–61 (2006).
- Nasmyth, K. & Haering, C. H. The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.* **74**, 595–648 (2005).
- Yu, H. G. & Koshland, D. Chromosome morphogenesis: Condensin-dependent cohesin removal during meiosis. *Cell* **123**, 397–407 (2005).
- Alexandru, G., Uhlmann, F., Mechtler, K., Poupard, M. A. & Nasmyth, K. Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105**, 459–472 (2001).
- Lee, B. H. & Amon, A. Role of Polo-like kinase Cdc5 in programming meiosis I chromosome segregation. *Science* **300**, 482–486 (2003).

15. Clyne, R. K. *et al.* Polo-like kinase Cdc5 promotes chiasmata formation and cosegregation of sister centromeres at meiosis I. *Nature Cell Biol.* **5**, 480–485 (2003).
16. Hornig, N. C. & Uhlmann, F. Preferential cleavage of chromatin-bound cohesin after targeted phosphorylation by Polo-like kinase. *EMBO J.* **23**, 3144–3153 (2004).
17. Klein, F. *et al.* A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell* **98**, 91–103 (1999).
18. Buonomo, S. B. *et al.* Disjunction of homologous chromosomes in meiosis I depends on proteolytic cleavage of the meiotic cohesin Rec8 by Separase. *Cell* **103**, 387–398 (2000).
19. Keeney, S., Giroux, C. N. & Kleckner, N. Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384 (1997).
20. Bergerat, A. *et al.* An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386**, 414–417 (1997).
21. Toth, A. *et al.* Functional genomics identifies monopolin: A kinetochore protein required for segregation of homologs during meiosis I. *Cell* **103**, 1155–1168 (2000).
22. Alani, E., Padmore, R. & Kleckner, N. Analysis of wild-type and *rad50* mutants of yeast suggests an intimate relationship between meiotic chromosome synapsis and recombination. *Cell* **61**, 419–436 (1990).
23. Bishop, D. K., Park, D., Xu, L. & Kleckner, N. Dmc1: A meiosis-specific yeast homolog of *E. coli* RecA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **69**, 439–456 (1992).
24. Shonn, M. A., Murray, A. L. & Murray, A. W. Spindle checkpoint component Mad2 contributes to biorientation of homologous chromosomes. *Curr. Biol.* **13**, 1979–1984 (2003).
25. Shonn, M. A., McCarroll, R. & Murray, A. W. Requirement of the spindle checkpoint for proper chromosome segregation in budding yeast meiosis. *Science* **289**, 300–303 (2000).
26. Cohen-Fix, O., Peters, J. M., Kirschner, M. W. & Koshland, D. Anaphase initiation in *Saccharomyces cerevisiae* is controlled by the APC-dependent degradation of the anaphase inhibitor Pds1p. *Genes Dev.* **10**, 3081–3093 (1996).
27. Nairz, K. & Klein, F. *mre11S*—a yeast mutation that blocks double-strand-break processing and permits nonhomologous synapsis in meiosis. *Genes Dev.* **11**, 2272–2290 (1997).
28. Visintin, R. *et al.* The phosphatase Cdc14 triggers mitotic exit by reversal of CDK-dependent phosphorylation. *Mol. Cell* **2**, 709–718 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank F. Lewitter for help with the Polo kinase substrate consensus sequence analysis, L.-S. Ee for technical assistance, K. Nasmyth and Y. Watanabe for communication of results before publication and B. Lee, A. Marston and members of the Amon Lab for input and critical reading of the manuscript. A.A. is an Investigator of the Howard Hughes Medical Institute. This research was supported by an NIH grant to A.A. and a NSF pre-doctoral fellowship to G.A.B.

Author Contributions G.A.B., B.M.K. and A.A. participated in the design of experiments. G.A.B. and B.M.K. performed all non-MS experiments. Y.Z., J.-E.K. and F.W. designed and performed all MS analysis and analysed MS results.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.A. (angelika@mit.edu).

Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways

Dirk Grimm¹, Konrad L. Streetz^{1†}, Catherine L. Jopling², Theresa A. Storm¹, Kusum Pandey¹, Corrine R. Davis³, Patricia Marion⁴, Felix Salazar⁴ & Mark A. Kay¹

RNA interference (RNAi) is a universal and evolutionarily conserved phenomenon of post-transcriptional gene silencing by means of sequence-specific mRNA degradation, triggered by small double-stranded RNAs^{1,2}. Because this mechanism can be efficiently induced *in vivo* by expressing target-complementary short hairpin RNA (shRNA) from non-viral and viral vectors, RNAi is attractive for functional genomics and human therapeutics^{3,4}. Here we systematically investigate the long-term effects of sustained high-level shRNA expression in livers of adult mice. Robust shRNA expression in all the hepatocytes after intravenous infusion was achieved with an optimized shRNA delivery vector based on duplex-DNA-containing adeno-associated virus type 8 (AAV8). An evaluation of 49 distinct AAV/shRNA vectors, unique in length and sequence and directed against six targets, showed that 36 resulted in dose-dependent liver injury, with 23 ultimately causing death. Morbidity was associated with the downregulation of liver-derived microRNAs (miRNAs), indicating possible competition of the latter with shRNAs for limiting cellular factors required for the processing of various small RNAs. *In vitro* and *in vivo* shRNA transfection studies implied that one such factor, shared by the shRNA/miRNA pathways and readily saturated, is the nuclear karyopherin exportin-5. Our findings have fundamental consequences for future RNAi-based strategies in animals and humans, because controlling intracellular shRNA expression levels will be imperative. However, the risk of oversaturating endogenous small RNA pathways can be minimized by optimizing shRNA dose and sequence, as exemplified here by our report of persistent and therapeutic RNAi against human hepatitis B virus *in vivo*.

To allow simple, efficient and persistent shRNA expression in entire mouse liver, we constructed a vector from AAV8. It carried a 'stabilized double-stranded' (sds) DNA genome, flanked by packaging signals derived from two heterologous AAV genotypes (see Supplementary Information)^{5–7}. We evaluated liver-directed shRNA expression by co-injecting firefly luciferase-expressing sdsAAV8 into mice together with five individual vectors encoding distinct anti-luciferase shRNAs (Supplementary Table 1). Complete liver transduction was achieved with doses as low as 2×10^{11} particles. Comparisons of bioluminescence in shRNA-injected and control animals showed efficient luciferase knockdown with two vectors for more than one month after injection (Supplementary Fig. 1).

Several mice treated with high doses of shRNA vector developed signs of severe toxicity (for example weight loss) and ultimately died within one month. A transient increase in luciferase expression in the liver of moribund mice indicated that acute events occurred in this organ (Supplementary Fig. 1). Numerous controls proved that toxicity was shRNA-specific and not related to vector components (not shown).

To confirm these findings we treated mice transgenic for the human α -1 antitrypsin (*hAAT*) gene with a series of anti-*hAAT* shRNA-expressing sdsAAV8 (Supplementary Table 1; representative results in Fig. 1a). Again, several shRNAs caused toxicity and morbidity, particularly the 25-mer at high doses, usually killing treated animals within one month. In most other mice RNAi was transient and reached a peak after two weeks, before *hAAT* levels rose sharply, often to surpass initial values. A notable exception was the low dose of 19-mer, which gave efficient and persistent *hAAT* knockdown for more than one year (Fig. 1a).

The data overall indicated that *in vivo* toxicity was frequent and not restricted to a particular shRNA or target. For further confirmation we created a battery of sdsAAV8 vectors encoding 40 additional distinct shRNAs (total 49; Supplementary Table 1). Infusion of high particle doses (10^{12} , ensuring complete liver transduction) into wild-type C57/BL6 or FVB mice ($n \geq 3$ each; both strains gave identical results) revealed that 36 of 49 shRNAs (73%) were severely toxic. Nearly half of all vectors (23 of 49; 47%) killed treated mice within two months. We concluded that lethality did not require the presence of an shRNA target (see also Supplementary Fig. 2).

Pathologically, the moribund mice often had severe ascites and widespread subcutaneous oedema. Gross and histological analyses of all major organs revealed that the main changes were in liver (Fig. 1b), with chronic hepatopathy characterized by lobular collapse, multifocal hepatocyte necrosis and an apparent regenerative effort demonstrated by marked anisocytosis and presence of numerous megalocytes (Fig. 1c). Additional signs of liver failure included elevated total and direct bilirubin levels (1.2–5.6 mg dl⁻¹, normal 0.65–0.85; direct bilirubin more than 85% of total), correlating with icteric sera. The prominent change in blood parameters was a decrease in total protein, albumin and globulin to nearly undetectable levels (less than 0.5 g dl⁻¹), probably causing the oedema and ascites. Liver enzymes were typically elevated in moribund mice at about one month after vector administration (alanine aminotransferase: 232–2665 IU l⁻¹, normal 76–160; aspartate aminotransferase: 444–5723 IU l⁻¹, normal 192–388; alkaline phosphatase: 192–894 IU l⁻¹, normal 171–183). All other parameters, including complete blood counts and electrolyte panels, were normal. Together with the fact that more than 90% of AAV vector DNA resided in the liver (not shown), these findings indicated shRNA-induced liver dysfunction as the cause of morbidity.

We found no evidence in ailing mice for activation of the RNA-dependent protein kinase/interferon pathway (previously controversially discussed as an adverse effect of shRNA expression⁸) or for cell cycle perturbation. ELISA-based screening of 45 sera (25 from ailing mice, 20 from healthy) showed normal interferon- $\alpha/\beta/\gamma$, tumour necrosis factor- α and interleukin-6 levels (not shown), and

¹Stanford University, School of Medicine, Departments of Pediatrics and Genetics, ²Department of Microbiology and Immunology, and ³Department of Comparative Medicine, 300 Pasteur Drive, Stanford, California 94305, USA. ⁴Hepadnavirus Testing, Inc., 331H Sierra Vista, Mountain View, California 94043, USA. [†]Present address: Department of Medicine III, University-Hospital Aachen, Pauwelsstrasse 30, 52074 Aachen, Germany.

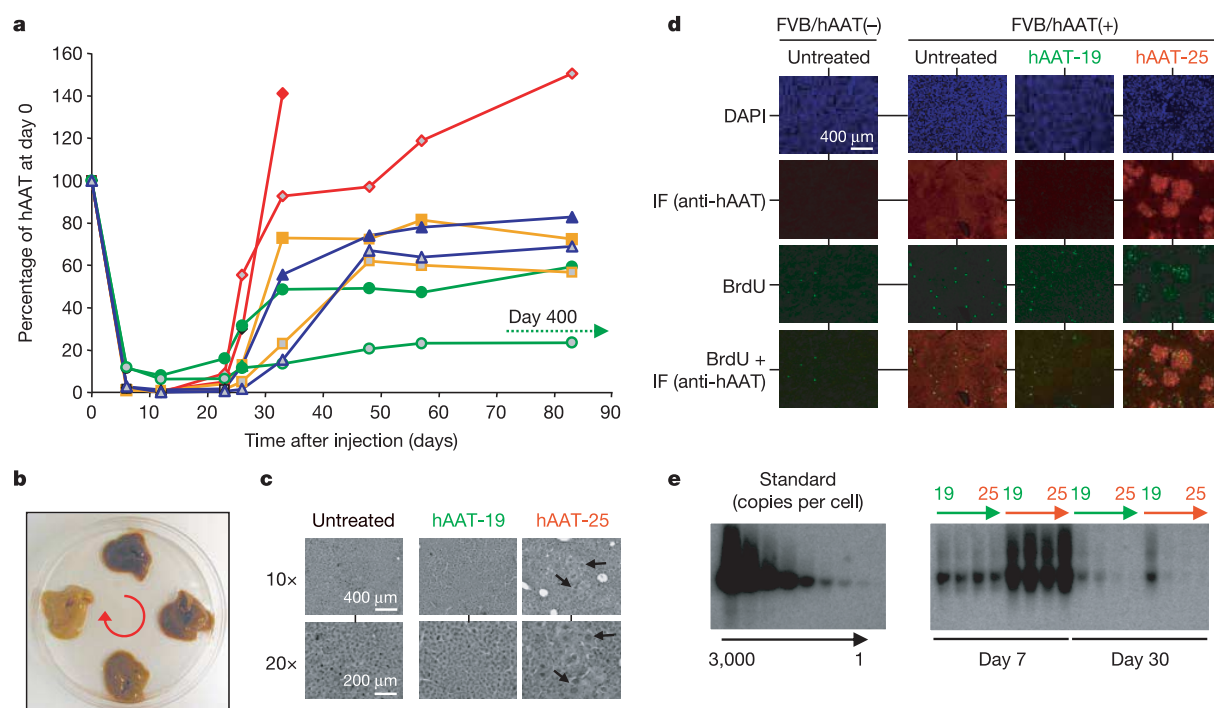


Figure 1 | Hepatocellular toxicity and liver regeneration from shRNA overexpression. **a**, hAAT-transgenic mice ($n = 6$) were injected with 10^{11} (low, filled symbols) or 10^{12} (high, open symbols) anti-hAAT shRNA-expressing sdsAAV8. Means of hAAT serum levels are relative to day 0 (error bars are omitted for clarity). Red, 25-mer; blue, 23-mer; orange, 21-mer; green, 19-mer. **b**, **c**, Transient RNAi coincided with liver toxicity, evident from abnormal morphology (**b**; arrow indicates increasing liver damage over

time) and histology (**c**; arrows indicate megakaryocytes, a sign of liver regeneration). **d**, Stable silencing with the 19-mer and abundant hAAT/bromodeoxyuridine (BrdU)-positive liver nodules with the 25-mer. DAPI, 4,6-diamidino-2-phenylindole; IF, immunofluorescence. **e**, Southern blot showing that hepatocyte death results in the loss of AAV vector DNA. Numbers indicate shRNA stem lengths; arrow colours show vector doses (green, low; red, high; see **a**).

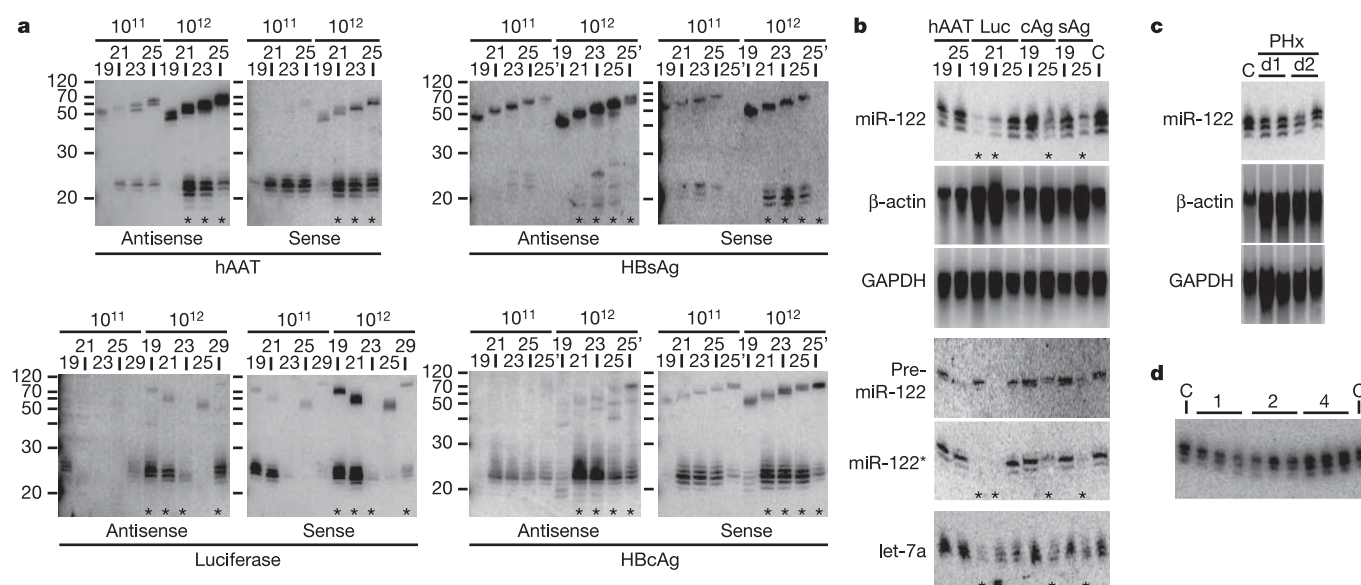


Figure 2 | shRNA overexpression saturates cellular miRNA pathways. **a**, Northern blot shows dose-dependent shRNA expression from a representative subset of vectors (Supplementary Table 1). Stem lengths are indicated above each lane, and marker sizes (in nucleotides) at the left. Asterisks indicate toxic shRNAs. **b**, Decrease in hepatocellular *miR-122* levels (asterisks) in mice overexpressing toxic shRNAs (northern blot with representative samples). The hAAT-25 mouse exemplifies rare recovery and return to normal *miR-122* levels. **c**, After partial hepatectomies (PHx) to

induce cell division⁹, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β -actin mRNA were typically elevated for two days, whereas *miR-122* levels remained stable. **d**, Continuous treatment with carbon tetrachloride to induce liver damage (demonstrated by elevated alanine aminotransferase (573–786 IU l⁻¹), icteric sera and abnormal morphology (not shown)) did not affect *miR-122*. Each lane represents one mouse per time point (weeks after first injection). Lanes marked C represent untreated control FVB mice.

western blots confirmed stable p21/p53 levels and a lack of eIF2- α phosphorylation (Supplementary Fig. 3). Although it was technically impossible to rule out off-targeting⁸, this seemed unlikely because 23 different shRNAs gave a similar lethal phenotype.

We proposed that overexpressing shRNAs led to hepatocyte death, often severe enough to cause liver failure and morbidity. We assume that in mice that escaped morbidity but showed transient RNAi, the intrinsic ability of liver to regenerate itself⁹ allowed dying hepatocytes to be replaced with new cells lacking shRNA. Direct proof for this idea was obtained by injecting hAAT-transgenic mice to induce either persistent hAAT silencing (19-mer, low dose) or toxicity (25-mer, high dose). When the latter seemed morbid and RNAi became transient (day 26), all livers were collected and stained (Fig. 1d). As predicted, mice treated with the 19-mer showed no hAAT expression, whereas the 25-mer group displayed manifold liver nodules of hAAT-positive (and thus shRNA-negative) hepatocytes, identified as actively dividing cells repopulating the liver by bromodeoxyuridine co-staining. This was corroborated by quantification of AAV vector DNA in livers early or late after injection: only the vector expressing the 19-mer persisted in the long term, whereas all other AAV genomes were almost completely lost during the process of liver damage and repopulation (Fig. 1e).

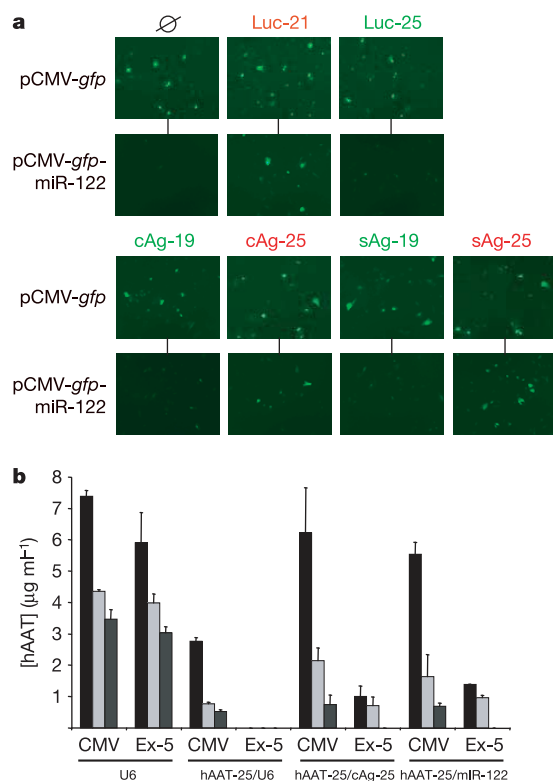


Figure 3 | shRNA overexpression inhibits *miR-122* functionality in liver cells. **a**, Human Huh7 liver cells were co-transfected with the shown shRNA plasmids (green, non-toxic; red, toxic; ϕ , empty plasmid control) and a *gfp*-encoding vector with or without a *miR-122*-binding site¹². GFP expression from the fusion construct was marginal in the presence of no or non-toxic shRNAs, but elevated with toxic shRNAs, indicating possible inhibition of *miR-122* function (up to 50% positive cells, in comparison with non-fused *gfp* controls). **b**, Wild-type FVB mice ($n = 3$ per group) were hydrodynamically transfected with 8 μ g of a Rous sarcoma virus (RSV)-promoter-driven hAAT gene and 12 μ g of a human U6 promoter plasmid (lacking shRNA), or 2 μ g of the hAAT-25 and 10 μ g of the empty U6, the cAg-25, or a *miR-122* plasmid. In addition, all mice received 5 μ g of a cytomegalovirus (CMV)-promoter-driven exportin-5 gene or an empty CMV control plasmid (pBCKMV; Stratagene). Plasma hAAT levels (means $\pm$ s.d.) were determined 1, 2 or 3 days after transfection (black, light grey and dark grey bars, respectively).

To understand the differences between various shRNAs, we studied their expression in injected mice. The shRNA precursors occasionally dominated the mature shRNAs, probably as a result of inefficient processing. Moreover, shRNA levels sometimes varied drastically between molecules differing in only two nucleotides. These findings might explain some historical controversies⁸ and highlight the necessity of experimentally controlling for shRNA expression. We frequently found a direct correlation of high shRNA levels with toxicity (Fig. 2a), indicating that a possible reason for hepatocyte death was oversaturation of the endogenous shRNA processing machinery. The latter is shared by miRNAs, cellular RNAs about 22 nucleotides long with central regulatory functions in mRNA translation, the cell cycle and development, for example, but also in cancer or virus replication^{10–12}. We therefore speculated that highly expressed ‘toxic’ shRNAs competed with miRNAs for intracellular processing, to such an extent that affected cells died.

Further RNA analyses supported this model, because toxicity and morbidity were correlated with decreased (by 54–81%) levels of the miRNAs *miR-122* (representing about 70% of all liver miRNAs¹³) or *let-7a* (Fig. 2b). Typically for repopulating liver, β -actin transcription was greater (unlike that of other cellular RNAs; Supplementary Fig. 4). When we induced liver regeneration by surgery (Fig. 2c) or chemical injury (Fig. 2d) we found that miRNA inhibition specifically resulted from shRNA overexpression and not from general liver injury. The correlation between toxicity and premature miRNAs was less clear, although the latter were frequently also present in lower concentrations like the processed forms (Fig. 2b, and not shown).

To assess whether shRNA overexpression compromised miRNA functionality, we co-transfected Huh7 cells (abundantly expressing *miR-122* (ref. 12)) with a *gfp* reporter fused to a *miR-122* target site, together with various shRNAs (Fig. 3a). Expression of green

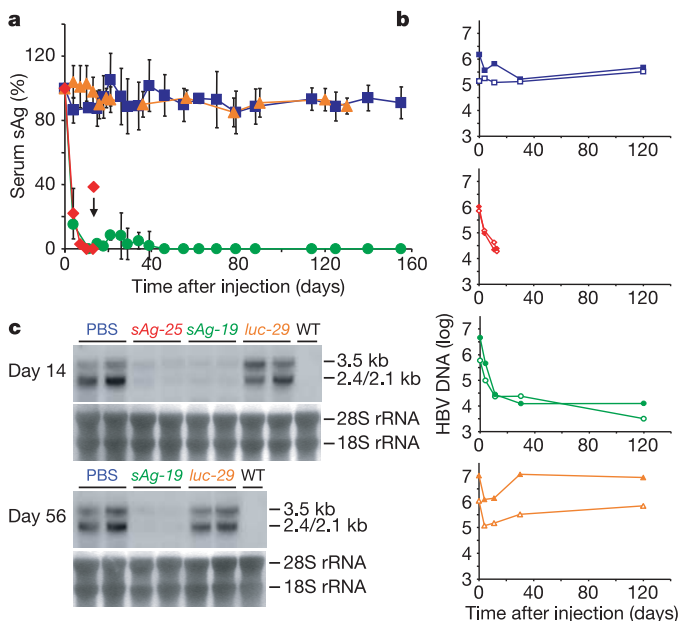


Figure 4 | AAV/shRNA-mediated persistent HBV inhibition in transgenic mice. **a**, HBV-transgenic mice ($n = 6$) were treated with anti-HBV (or anti-luciferase control (orange)) shRNA-expressing sdsAAV8 (3×10^{11} particles per mouse). HBsAg knockdown is shown as a percentage of day 0 (means $\pm$ s.d.). The 25-mer anti-HBV shRNA (red) caused early death, whereas the 19-mer (green) persistently and efficiently repressed HBsAg. Blue, PBS. **b**, Serum HBV DNA level (virus load) confirms long-term HBV knockdown *in vivo*. Colours as in **a**. **c**, Northern blots show nearly complete and persistent elimination of viral mRNAs in mice treated with the non-toxic 19-mer. No effect was seen in negative controls. Shown are two representative mice per group. WT, HBV-negative wild-type FVB mouse; kb, kilobases.

fluorescent protein (GFP) was marginal in the presence of a non-lethal shRNA or no shRNA, because of efficient *miR-122*-mediated cleavage of the fusion mRNA¹². In contrast, co-delivery of highly expressed, toxic shRNAs inhibited *miR-122* processing, demonstrated by greater expression of GFP. These results corroborate a recent study *in vitro*¹⁴.

To assess shRNA/miRNA competition *in vivo*, we performed hydrodynamic plasmid injections in wild-type FVB mice (Fig. 3b). We co-expressed the *hAAT* gene with a limiting amount of the anti-*hAAT* 25-mer shRNA and an excess of the toxic *cAg-25* shRNA or *miR-122*. As expected, *hAAT-25* shRNA functionality was reduced in the presence of *cAg-25* or *miR-122* in comparison with controls without competitor.

The nuclear karyopherin exportin-5 was identified as an essential and shared component of the shRNA and miRNA pathways^{15–17}. We therefore overexpressed exportin-5 in the transfected mice and found that it enhanced *hAAT* silencing, in the absence or presence of shRNA/miRNA competitors. The finding of exportin-5 as a limiting factor in the RNAi pathway is consistent with our observation that toxicity did not require an shRNA target, hinting at saturation of one or more early components. It also corroborates the current theory that the main function of exportin-5 is the nuclear export and stabilization of shRNAs/miRNAs, and not other critical macromolecules involved in translation, because we did not observe a global protein synthesis shutdown in ailing mice (see also Supplementary Fig. 2). To conclusively unravel all the molecules involved, future work will have to focus on selective inhibition or overexpression of other factors shared by shRNAs/miRNAs, such as Dicer or other parts of the RNA-induced silencing complex¹⁸. Our sdsAAV8/RNAi technology will prove useful for such approaches.

If the pathway can be oversaturated, efficient and persistent RNAi *in vivo* should be possible with the use of minimal vector doses and/or inherently weakly expressed shRNAs. We tested this in a therapeutically relevant model of AAV/RNAi-mediated gene silencing, by using hepatitis B virus (HBV).

We injected mice, transgenic with the full-length HBV genome and producing viral particles¹⁹, with low doses of sdsAAV8 expressing the non-toxic *sAg-19* shRNA. A lethality control group received the *sAg-25* vector, whereas *luc-29* (Supplementary Table 1) sdsAAV8 (non-lethal at this dose) or PBS served as negative controls. Blood was collected to quantify serum HBsAg and viral loads, and livers were harvested for RNA analyses.

Within two weeks, HBsAg decreased rapidly to below 10% of starting levels with both anti-HBV vectors, before *sAg-25*-treated mice died (Fig. 4a). The *sAg-19* group showed no signs of morbidity, and HBsAg stabilized six weeks after injection at nearly undetectable levels for the entire duration of the experiment (more than 5 months). Efficient HBV suppression was corroborated by quantification of viral loads (Fig. 4b) and transcripts (Fig. 4c). Both anti-HBV vectors caused a more than 32-fold decrease in serum viral DNA within two weeks, and at four months the decrease with the *sAg-19* shRNA stabilized at 148 ± 3 -fold ($n = 3$). HBV mRNA had decreased in amount by more than 90% two weeks after *sAg-19* injection relative to controls, and was barely traceable at week 8.

There is increasing evidence that introducing siRNAs into, or expressing shRNAs in, mammalian cells can have non-specific effects, including off-target silencing and activation of the interferon system⁸. However, here we have linked *in vivo* shRNA overexpression to tissue damage and frequent lethality in adult animals. We detected an intracellular shRNA threshold for toxicity, causally linked to fatal competition with endogenous miRNA processing and functionality. On the basis of this and published data^{14,15}, we speculate that the competition involves the overloading of exportin-5 (whose main function is nuclear export and stabilization of shRNAs/miRNAs) and perhaps other common factors (Supplementary Fig. 5). Individual miRNAs were recently blocked in mice by using 'antagomirs'²⁰. The authors found minor non-lethal phenotypes, substantiating our

model that shRNA-induced toxicity requires global miRNA inhibition, with *miR-122* and *let-7a* merely being examples. Additional support comes from studies with siRNAs to trigger silencing *in vivo*, in which lethality was also absent (although it is uncertain whether the same high siRNA concentrations were achieved)⁸. We can explain this by the fact that siRNAs enter the RNAi pathway at a late step²¹, thus avoiding the saturation of exportin-5 or other early factors. Last, our competition model could explain why efforts to establish transgenic mice expressing multicopy shRNA sequences at high levels have remained unsuccessful²².

We believe that by identifying the risk of unintentional miRNA inhibition by shRNA overexpression, our study will help to pave the way for future *in vivo* RNAi applications. We believe that monitoring and controlling intracellular shRNA levels is imperative for guaranteeing stable *in vivo* gene silencing while mitigating adverse effects. Fortunately, there are several options, such as shRNA expression from inducible or tissue-specific RNA polymerase II promoters^{23,24}. Until those advances are in place, the present and previous work^{25–28} show that, under optimized conditions, the current generation of shRNA vectors already permits an effective and stable *in vivo* silencing of endogenous and exogenous genes. This fuels hope that the wide use of RNAi for functional genomics, or as a new therapeutic in humans, will become a reality.

METHODS

See refs 29, 30 and Supplementary Information for more details.

ShRNAs were cloned into the sdsAAV vector and were expressed from the human U6 or H1 promoter. MicroRNA cleavage was measured by transfecting a plasmid containing a *gfp* gene with a *miR-122* site into Huh7 hepatoma cells; co-transfection with the shRNA expressing plasmids resulted in an enhanced fluorescence signal when miRNA-mediated cleavage was inhibited. Experiments *in vivo* were performed in FVB or C57/BL6 mice.

Received 26 September 2005; accepted 7 April 2006.

- Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
- Novina, C. D. & Sharp, P. A. The RNAi revolution. *Nature* **430**, 161–164 (2004).
- Dykxhoorn, D. M. & Lieberman, J. The silent revolution: RNA interference as basic biology, research tool, and therapeutic. *Annu. Rev. Med.* **56**, 401–423 (2005).
- Hannon, G. J. & Rossi, J. J. Unlocking the potential of the human genome with RNA interference. *Nature* **431**, 371–378 (2004).
- Nakai, H. *et al.* Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice. *J. Virol.* **79**, 214–224 (2005).
- McCarty, D. M., Monahan, P. E. & Samulski, R. J. Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene Ther.* **8**, 1248–1254 (2001).
- Grimm, D. & Kay, M. A. From virus evolution to vector revolution: use of naturally occurring serotypes of adeno-associated virus (AAV) as novel vectors for human gene therapy. *Curr. Gene Ther.* **3**, 281–304 (2003).
- Jackson, A. L. & Linsley, P. S. Noise amidst the silence: off-target effects of siRNAs? *Trends Genet.* **20**, 521–524 (2004).
- Nakai, H. *et al.* Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction *in vivo*. *J. Virol.* **75**, 6969–6976 (2001).
- Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
- Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
- Jopling, C. L., Yi, M., Lancaster, A. M., Lemon, S. M. & Sarnow, P. Modulation of hepatitis C virus RNA abundance by a liver-specific microRNA. *Science* **309**, 1577–1581 (2005).
- Lagos-Quintana, M. *et al.* Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**, 735–739 (2002).
- Yi, R., Doehle, B. P., Qin, Y., Macara, I. G. & Cullen, B. R. Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA* **11**, 220–226 (2005).
- Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011–3016 (2003).
- Bohnsack, M. T., Czapinski, K. & Görlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA* **10**, 185–191 (2004).

17. Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. & Kutay, U. Nuclear export of microRNA precursors. *Science* **303**, 95–98 (2004).
18. Sontheimer, E. J. Assembly and function of RNA silencing complexes. *Nature Rev. Mol. Cell Biol.* **6**, 127–138 (2005).
19. Marion, P. L., et al. in *Frontiers in Viral Hepatitis* (eds Schinazi, R. F., Rice, C. & Sommadossi, J.-P.) 197–209 (Elsevier Science, Amsterdam, 2002).
20. Krutzfeldt, J. et al. Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**, 685–689 (2005).
21. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343–349 (2004).
22. Cao, W., Hunter, R., Strnatka, D., McQueen, C. A. & Erickson, R. P. DNA constructs designed to produce short hairpin, interfering RNAs in transgenic mice sometimes show early lethality and an interferon response. *J. Appl. Genet.* **46**, 217–225 (2005).
23. Xia, H., Mao, Q., Paulson, H. L. & Davidson, B. L. siRNA-mediated gene silencing *in vitro* and *in vivo*. *Nature Biotechnol.* **20**, 1006–1010 (2002).
24. Ventura, A. et al. Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* **101**, 10380–10385 (2004).
25. Hommel, J. D., Sears, R. M., Georgescu, D., Simmons, D. L. & DiLeone, R. J. Local gene knockdown in the brain using viral-mediated RNA interference. *Nature Med.* **9**, 1539–1544 (2003).
26. Xia, H. et al. RNAi suppresses polyglutamine-induced neurodegeneration in a model of spinocerebellar ataxia. *Nature Med.* **10**, 816–820 (2004).
27. McCaffrey, A. P. et al. Inhibition of hepatitis B virus in mice by RNA interference. *Nature Biotechnol.* **21**, 639–644 (2003).
28. McCaffrey, A. P. et al. RNA interference in adult mice. *Nature* **418**, 38–39 (2002).
29. Grimm, D. Production methods for gene transfer vectors based on adeno-associated virus serotypes. *Methods* **28**, 146–157 (2002).
30. Jenkins, D. D. et al. Donor-derived, liver-specific protein expression after bone marrow transplantation. *Transplantation* **78**, 530–536 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Wilson for providing the AAV8 packaging plasmid, P. Sarnow for the *gfp* fusion and miR-122 expression plasmids, H. Doege and A. Stahl for the H1-driven shRNA cassettes, I. Macara for the exportin-5 vector, D. Haussecker and B. Garrison for critically reading the manuscript, and J. S. Lee and H. Xu for technical assistance. This work was supported by grants from the National Institutes of Health (to M.A.K.) and the Anna Ng Charitable Foundation (to M.A.K.).

Author contributions D.G. performed and designed (with M.A.K.) most of the included studies. K.L.S. maintained the hAAT-transgenic mice and performed all injections and histological liver analyses. C.L.J. performed crucial steps of the small RNA Northern blot analyses and provided the *gfp* plasmids used in Fig. 4a. T.A.S. and K.P. generated the sdsAAV8 preparations and helped with the DNA, RNA and protein analyses. C.R.D. performed the complete mouse pathologies. P.M. and F.S. provided and maintained the HBV-transgenic mice. M.A.K. supervised the research project, and assisted in the experimental design. All authors discussed the experimental results and had input into the writing of the final manuscript.

Author Information Reprints and permissions information is available at www.npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.K. (markay@stanford.edu).

naturejobs

**THE CAREERS
MAGAZINE FOR
SCIENTISTS**

Next month a group of students will launch their bid to set up a national association for postdocs in Britain (see page 546). A worthy plan — although it does raise the question of why this hasn't been done before. The most likely reason is that starting a postdoc organization means taking time away from work. It isn't easy trying to run an organization to improve your working conditions and prospects when the time it takes can have a detrimental effect on the very things you're trying to make better. Fortunately, creating such an organization need not be difficult, especially as both the European Union and the United States have established models in the shape of EURODOC and the US National Postdoctoral Association (NPA).

Why are these organizations so important? Because without them, myriad postdoc complaints — such as low stipends, a lack of benefits and a dearth of career advice — go, if not unheard, then largely ignored. Having a national presence gives a voice and a face to these concerns.

But the key to better treatment for UK postdocs lies not just with the fledgling national organization — it also rests in the hands of smaller organizations at each of the country's research institutions. As part of its success, the NPA has encouraged the foundation of local postdoc organizations and offices (see *Nature* **441**, 249; 2006). The fledgling UK association would do well to follow this model. National organizations provide a unified voice and can spearhead drives to collect data on salaries, benefits and long-term career prospects. But it is the local organizations that can address specific problems at individual institutions and help individual postdocs feel less isolated.

Yes, establishing these local organizations takes time. But thanks to the NPA and EURODOC, fellows in Britain or any other country don't need to reinvent the wheel. Rising to the challenge will prove to be more than worthwhile — especially if it means that students have better postdoc experiences and find it easier to make the transition to permanent employment.

Paul Smaglik, *Naturejobs* editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Gene Russo

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
Peter Bless
Classified Sales Representatives
Tel: +1 800 989 7718
East USA/Canada: Andrew Bennie
**NIH/Maryland/New York/
Pennsylvania:** Shelley Cohen

San Francisco Office
Classified Sales Representative:

Michaela Bjorkman
West USA/West Corp. Canada
225 Bush Street, Suite 1453
San Francisco, CA 94104
Tel: +1 415 781 3803
Fax: +1 415 781 3805
e-mail: m.bjorkman@naturesf.com

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
Advertising Production Manager:
Stephen Russell
To send materials use London
address above.
Tel: +44 (0) 20 7843 4816

Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
Naturejobs web development:
Tom Hancock
Naturejobs online production:
Catherine Alexander

European Satellite Office
Patrick Phelan
e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku,
Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
e-mail: a.watanabe@natureasia.com

Animal intelligence

Use of animals for testing early in the drug-development process aims to provide vital information to make new drugs safe and effective — and the process is being constantly refined. **Hannah Hoag** finds out what is involved.

The decision to carry a drug through to human clinical trials depends on many factors, but safety tops that list. Scientists abandon a huge number of molecules early in the pipeline because of safety concerns — only about 0.1% of compounds that enter preclinical testing move on to clinical trials. The animal-testing phase is a critical piece of the preclinical development of any medicine.

The recent TGN1412 trial, which provoked an extreme reaction in the subjects taking the drug, shows that even now that process isn't a safe bet (see *Nature* **440**, 855–856; 2006). And the highly publicized withdrawal of a handful of blockbuster drugs from the market in recent years has pushed pharmaceutical companies to introduce safety tests earlier in development and to devise novel tests that will allow scientists to halt drugs sooner.

Traditionally, animal-safety studies are done just before clinical trials. The studies reveal a drug candidate's features and its action in the body, but they are costly and labour-intensive. Recently, in an attempt to save time and money, the drug industry has begun to conduct animal studies during the very earliest stages of drug discovery. By doing this, the companies hope to eliminate poor or unsafe compounds earlier in the process. These studies use far fewer animals, or even unconventional drug-discovery animal models such as zebrafish or roundworms, to help scientists decide which compounds to promote to the development stage.

Taking a lead

The involvement of a pharmacologist or biologist with a toxicology background usually starts during 'lead identification', when drug candidates are being picked out, and 'lead optimization', when they're being tweaked (see *Nature* **439**, 886–877; 2006). Early safety studies in animals are becoming more routine. "It gives a discovery scientist the confidence to either take the molecule forward in the discovery chain or drop it and prioritize other molecules," says Geeta Sharma, associate director for biology at GVK Bio, a contract organization in Hyderabad, India, that provides chemical, biological and clinical services.

During these early stages, scientists concerned about a drug candidate's pharmacokinetics — its absorption, distribution, metabolism and elimination — work closely with medicinal chemists to find the right molecules to pursue. These molecule-builders make a series of slightly different structures and send them for



Clear value: early tests aim to weed out harmful drugs.

safety evaluation in animals. The results are then sent back to the medicinal chemists so that they can tweak the molecules further. "The key thing is the rapidity with which the data are generated, which then cycles back to our colleagues and partners in medicinal chemistry and affects the next series of structures," says Michael Silber, head of safety and technical services at Roche's Palo Alto facility in California.

Next come formal animal studies, which are carried out under laboratory practices that comply with industry standards and follow regulatory agency guidelines. "These formal safety assessments form the basis of filing with regulatory agencies to show that the molecule is safe," says Kevin Stark, Amgen's senior director of strategic operations.

Such tests lay the groundwork for human clinical trials. The US Food and Drug Administration (FDA) requires data showing that "the drug is reasonably safe for use in initial, small-scale clinical studies". At the very least, a company must provide a pharmacological profile, including absorption, distribution, metabolism and elimination; determine acute toxicity in two animal species; and conduct short-term toxicity studies, usually lasting 28 days, to mimic human use. The term ADME-tox describes this battery of tests. The goal of ADME-tox testing is to maximize safety and efficacy predictions so as to minimize drug-development costs.

"We are trying to understand what the drug does to

CORBIS

normal animals, typically a rodent and non-rodent species, and making sure there is a therapeutic window," says Scott Biller, head of chemical discovery at Novartis. "It is probably the most important aspect of the preclinical development phase — to characterize the toxicity of your molecule relative to efficacy."

Drugs typically fail in the preclinical phase because of poor pharmacokinetics or toxicology. "Poor bioavailability, poor blood levels and too short a half-life are a major cause," says Jon Mirsalis, director of preclinical development at SRI International, a contract research organization in the business of discovering and developing new medicines. "Poor pharmacokinetics probably kills 60–70% of the drugs that die in preclinical. The other 30% die for toxicology reasons."

Quest for data

New tools, such as genomics, proteomics and the use of biomarkers (physical or molecular indicators of treatment efficacy or disease progression), promise to improve clinical success rates. "It has to do with cost, reducing animal use and getting better data early on. If you are confident that this drug is going to shut down enzyme X, you had better have some data that that is happening, and if you can't show that, you had better not move the drug forward," says Mirsalis.

Toxicogenomics continues to flourish as a predictive tool. Using tools such as microarrays, this relatively new genomics approach looks for changes in gene expression in live animals in the days following treatment, rather than waiting for the physiological and biochemical changes that might occur two to four weeks after treatment begins. The speed of testing helps identify compounds with toxic liabilities before too much time and resources have been invested.

Novel animal models that will detect failures earlier are slowly being introduced. In its Critical Path Initiative, the FDA called for new animal models to test the safety of novel drug candidates, estimating that a 10% improvement in predicting failures would save US\$100 million per drug in development costs.

One trend is for improving current animal models and adding to their complexity. But emerging *in vivo* models that use species such as roundworms, fruitflies and zebrafish are also gaining ground. Although these animals have long been a mainstay of biological sciences, they remain relatively novel to the pharmaceutical industry.

Zebrafish may provide opportunities to accelerate the process of drug discovery (L. I. Zon and R. T. Peterson *Nature Rev. Drug Disc.* 4, 35–44; 2005). Their tiny size,



Jon Mirsalis: looking for skills and leadership.

large numbers, rapid development and transparent bodies make them suited to high-throughput screening.

Although less established than traditional mammalian models, zebrafish are easier to use, faster and inexpensive, says Nina Sawczuk, co-founder and chief executive of Zygogen, a biotech company based in Atlanta developing zebrafish models of human disease and toxicity for use in drug discovery. "The drug companies we're talking to are very interested in using zebrafish for toxicity screening early to validate targets and prioritize compounds," she says. Novartis recently established a Models of Disease Center (located in Cambridge, Massachusetts, and Basel, Switzerland) to develop zebrafish models of diseases for pharmacological studies, alongside the company's development of mouse models.

How to stand out in a crowd

Whether it's drug candidates or job candidates, choosing those with the greatest potential depends on a variety of factors. Job candidates with a PhD from a world-class institution will stand out, but without relevant work experience offers will remain scarce. Interpersonal skills and evidence of leadership are as important. "There are those who have a 4.0 and graduate *summa cum laude*, but the résumé is blank. They are not what we are looking for," says Mirsalis.

Graduates in toxicology, pathology, biology and lab-animal science are popular. "They should have some knowledge of the behaviour of a molecule in living systems and the adverse effects it could cause," says Stark. Maths graduates may find entry-level jobs doing biostatistical programming for ADME-tox groups.

"Experience at a discovery company definitely helps," says Niranjan Reddy, director of human resources at GVK Bio. "For a junior scientist, it is more important to be able to handle animals and conduct experiments, whereas a senior position would require knowledge of a particular therapeutic area." The breakdown of experience depends on where your talents will be focused. Weeding out drug candidates during the early stages requires a better understanding of the disease area, whereas ADME-tox studies require a solid understanding of animal models of disease and handling animals, says Silber.

Scientists carrying out formal safety and toxicity studies typically have a higher degree and several years of experience, perhaps as a postdoc in a toxicology lab. Veterinary qualifications are also common among this group, says Ken Batchelor, senior vice-president of GlaxoSmithKline's centre of excellence for drug discovery, metabolic and viral diseases.

Hiring practices can be difficult to predict, but contract research organizations tend to fare well in a variety of economies — deCODE Chemistry in Woodridge, Illinois, recently doubled the size of its drug safety and metabolism team. Business at SRI has tripled in six years, boosting the preclinical development group from 35 to 95.

Elsewhere, both large and small drug companies are eager to fill spots in *in vivo* pharmacology, ADME and safety testing. "It's an extremely rewarding area for scientists to be working in, to select and identify the needle in the haystack that ultimately turns out to be an important new medicine," says Silber.

Hannah Hoag is a freelance writer based in Montreal, Canada.

Cheap and easy: zebrafish are gaining ground in toxicology labs.



M. ZHU

MOVERS

Jai Nagarkatti, president and chief executive, Sigma-Aldrich, St Louis, Missouri



2005–06: Chief operating officer, Sigma-Aldrich, St Louis, Missouri

2003–05: President, scientific research division, Sigma-Aldrich, St Louis, Missouri

2000–03: President of fine-chemical division, Sigma-Aldrich, St Louis, Missouri

1987–99: President, Aldrich, Milwaukee, Wisconsin

Jai Nagarkatti is an oddity in today's corporate world. He has spent his entire career at Sigma-Aldrich, a \$1.7-billion life-science and chemical company based in St Louis, Missouri. Captivated by chemistry at an early age, he left his native India for the United States to get a PhD in synthetic organic chemistry at Texas A&M University.

Throughout graduate school, Nagarkatti had ordered his chemicals from Aldrich, so it seemed natural to apply for a job there. Starting as a production chemist, he has risen through the ranks over the past three decades — a journey to his current role as president and chief executive that he says has gone fast.

Only last year did he receive his first formal business training, through Harvard Business School's Advanced Management programme. Nagarkatti's three months in the programme exposed him to new trends in organizational structure, employee development and, perhaps most importantly, strategic planning for a global economy. Learning primarily through case studies, Nagarkatti says the most important insights came from being able to interact with peers in related industries across the globe. "Not only do you learn about conducting business internationally, you get insight into different cultural approaches to problem-solving," he says.

In many ways, his lack of formal business training was advantageous. Although he intuitively understood asset management and capital funding, Nagarkatti says it is his focus on listening — to both customer needs and employee ideas — that remains the driving force of his success. "Business is common sense," he says.

Nagarkatti carries on a company tradition of cultivating talented staff, which he cites as his greatest achievement. Following the Harvard programme, he selected ten high-potential employees from different parts of the company to help him develop a strategic plan for long-term growth. The plan, implemented last year, includes expanding the company's presence in growing economies such as the Pacific Rim countries and building on its Internet presence.

Over the years, Nagarkatti has rebuffed offers to work for competitors or other industries. Early on, he did entertain the notion. But he preferred to show allegiance to a company that had treated him fairly and rewarded him accordingly. "The grass always looks greener on the other side until you are there," he says, adding that he couldn't walk away from the company he has helped build. ■
Virginia Gewin

SCIENTISTS & SOCIETIES

Britain's postdocs unite

The longer my postdoc takes, the more time I get to spend on the beach. My research involves collecting and analysing seaweed on the shores of Plymouth, in England's sunny southwest. But I eventually want to become an independent investigator and, as seaweed research is hardly 'big science' and my institution is small, there are few visible career options. No postdoc from my institute has landed a tenured position for over five years.

Only about a quarter of UK postdocs get those sought-after jobs, I've found, and there's a real shortage of information, especially for young investigators trying to establish themselves after postdocs.

Concerned about this, I set off to look for the British equivalent of the US National Postdoctoral Association (NPA), which pushes for better pay and benefits, and identifies problems through regular surveys. To my surprise, I discovered that no such organization exists in Britain.

Last year, I chatted with one of the NPA's executive officers, Chris Blagden, about how to start a UK association. On Chris's advice, I contacted colleagues at other institutions to form an ad hoc steering committee. Our efforts bore fruit this month, when the UK Research Councils agreed to fund an inaugural meeting of the UK National Research Staff Association at University College London on 22 June. About 30 postdocs

will attend, representing institutes from Sussex to Dundee and disciplines from computing to midwifery.

UK postdocs need a national voice. The government is under pressure to retain young scientists, to compete with the emerging economies of China and India, but no one has asked us what it would take to keep us working in science. Although our welfare state makes healthcare and benefit worries less acute than in the United States, postdocs are often treated as temporary technicians, not as trainee principal investigators. The resulting emphasis on short-term productivity means no clear career structure exists.

We aim to found an association to advance UK research by fostering the creativity and independence of postdocs. To do this we need to gather information about working conditions and aspirations, and bring together postdocs, mentors and potential employers at workshops and conferences.

Our trickiest objective may simply be to create a self-sustaining structure. But with a potential membership of some of the smartest and most motivated 25–35-year-olds in the country, we're confident that we can establish a lasting, productive organization. ■

John Bothwell is a postdoctoral fellow at the Marine Biological Association of the UK, Plymouth, UK.

GRADUATE JOURNAL

Clocking out

The watch that I wear belonged to my grandfather. It's the kind that you have to wind up periodically. When fully wound, it runs a bit fast. Towards the end of the day it runs a bit behind. But as long as I remember to wind it every night, it's pretty accurate on average. My graduate-school experience has run to a timetable every bit as erratic as my watch.

My first few years went by quickly. Classes came at me in double time. No sooner had I finished my fledgling experiments than my qualifying exam was upon me. My first few years of graduate school were one of those frantic black-and-white films from the 1920s. But the last year has been a silent art film: poignant and slower. My final few experiments seem to unfold frame-by-frame. Every minute at the lab bench makes graduation seem a more distant prospect.

I think of my grandfather often when I wind my watch. Doing so puts things in some perspective. My grandfather overcame much greater obstacles in war-time Holland. He took enormous pride in doing what he believed in and doing it well. Winding my watch reminds me, too, that time always seems to balance out in the end. The thrill and rush of starting a scientific project inevitably give way to a more measured period of working out the details. Although eager to graduate, I'm willing to trust that finishing a PhD I'll be proud of is well worth the extra time. ■

Milan de Vries is a molecular-biology graduate student at the Massachusetts Institute of Technology.

Paratext

Step into another world.

Scarlett Thomas

You now have one choice.

You...I've been teaching my students in my apartment once a month since the campus shut down its offline operation and joined the network permanently. While I talk to the students about the concept of the post-human, which they understand to be "Like, *dead?*", my pod whirrs away in the corner, connected to the network (of course) but more specifically to the pods of my 15 closest friends.

A feat of physics created these quantum pods about ten years ago: grey boxes that look just like twentieth-century Apples but are capable of things you couldn't even do in fiction then. Our connected pods are running an AI program that was banned in 2012 after some government drone finally got around to seeing *The Matrix*. It's set up so that if a pod becomes conscious it will call one of our new-but-vintage cell phones (our revolutionary props). We don't know what we'll do then, but it's unlikely to happen anyway on a combination of 15 QII processors. Pods as the new proletariat? Whatever.

Becky said that when (not if) the machines start to think, their thoughts will be in code. She said that if you think using the stuff your world is made of then you make your own world. Then Rex wrote a story that had a guy walking home in the rain as the network awakens and all the lights start to go out in the Games and the Q-cars and the shops. When the guy gets home and looks at his pod, the only thing on there is a screensaver showing a garden going on forever — whatever forever is on a machine.

Becky was like: "But what does a machine *want?*"

It's 3:45; the sky is orangey outside and my students shrug every time I try to teach them anything. I glance at the cell phone. My agent hasn't called for days and I think my new novel's been rejected. What can I tell the students today? Once upon a time there were...records? Books? Malls? Last

week we talked about retro and nostalgia — all the fun of the postmodern. I told them I'd once thought people would run out of ways of recycling only the hairstyles, expressions of disgust and styles of wearing denim that existed between the invention of television and its death. Not even wrong.

I probably count as retro: my jeans; my sneakers. I don't have nano-shit on my face, ready to twitch my eye make-up into Beyoncé circa 2006, or Lt Uhura circa 1969. I can't — don't want to — make my hair into ringlets at the touch of a button. I don't wear a layer of thermo

sky, and now there are dwarves, elves and halflings in the doorways of what used to be Marks & Spencer, Virgin Megastore and Build-a-Bear Workshop.

The Games have many authors but essentially one story: you're on a quest to save the world, and you and some 'friends' must pool your potions, weapons and armour to fight the dragons, trolls and deformed birds roaming underneath the old shopping centre. The world can be threatened in several different ways. Lately I've noticed it's the online, not the offline, world you're supposed to save, but the result of saving it is the same: you get shopping credits, and violent sex with

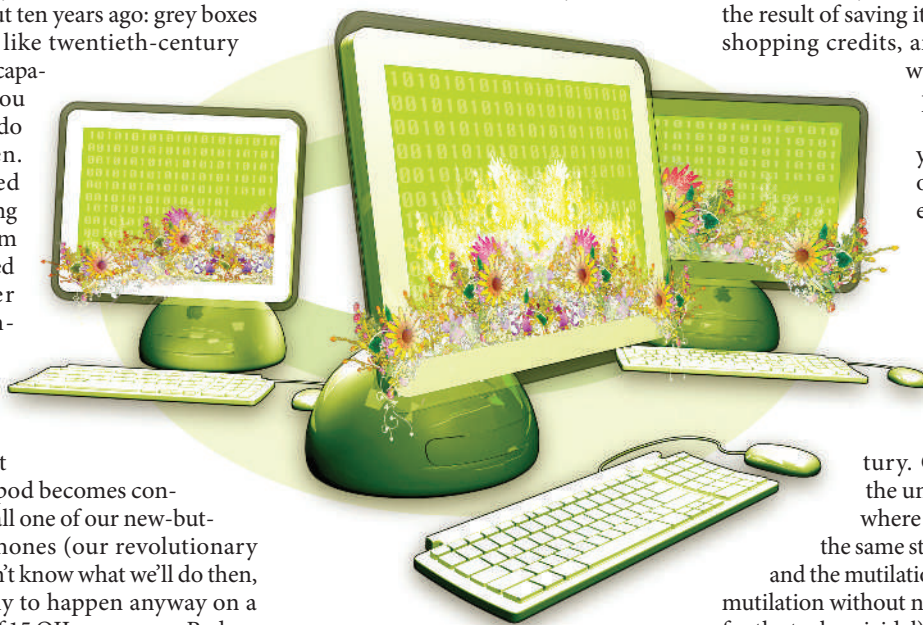
whoever was on the picture on the Q-display outside. But you can have violent, offline sex with whoever you want in the Games. That's the whole point. You can hate the Games, but you can't avoid them; they're like shopping malls at the turn of the century.

Of course, I prefer the underground versions where the story (it's always the same story) has some depth, and the mutilation is worse (although mutilation without nano-patching is only for the truly suicidal).

When my novel is rejected, I'll end up writing for the Games. I'll write about evil siblings, destiny and nemesis, and I still won't make any money. And the university will eventually get something like the old Microsoft Office Paperclip to teach my students. It's all so depressing; I might not go home yet. Maybe just one game...

I walk up to the old bookshop, hold up my wrist and then type Y into the console. My phone rings. I look at the display. It's not my agent; it's my pod. No. Seriously? I check again. I wait a few seconds for flickering lights and/or the end of the world. Nothing. Then I walk through the doorway of the bookshop and step onto the grass.

Scarlett Thomas is the author of the widely acclaimed novels *Going Out* and *PopCo* and the forthcoming *The End of Mr Y*. She lives in Canterbury, UK, and is currently working on her new novel, *Death of the Author*.



so thick that I hardly need clothes, like these kids: ready for the Games. Big surprises of the past 20 years? Obesity's gone, and no one plays videogames. Temporary cancers work wonders; and after governments, universities and shops all scrambled into binary, the Games finally went offline. So now there's real-time carnage amid the post-corporate concrete.

After class I walk through town to the only market, there for freaks like me who want occasional lettuces grown in fields, and real apples with pips. What has my pod ordered for me lately? Oh yes: a crate of stale ramen and two crates of non-vegetarian box-meals. At this time of day the Games are almost in Sleep mode, with their 24/7 neon pooling into the pale sky. The neon displays come in off the network, of course, as do the characters, stories and 'magic'. But when you play them you're off the network, and the pain is real. On my way back the neon has split from the